IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of   )
)
BEN SHEN, WEN LIU, STEVEN D.   )
CHRISTENSON and SCOTT STANDAGE   )
)
For: GENE CLUSTER FOR   )
        PRODUCTION OF THE ENEDIYNE   )
        ANTITUMOR ANTIBIOTIC C-1027   )
_____)   San Francisco, California

Patent Application
Assistant Commissioner for Patents
Washington, D.C. 20231

By Express Mail No: **EL160743652US**
Dated: January 5, 2000

PATENT APPLICATION TRANSMITTAL

Sir:

Transmitted herewith for filing is the patent application of inventor(s) Ben Shen, Wen Liu, Steven D. Christenson and Scott Standage, for "GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR ANTIBIOTIC C-1027." Enclosed are:

1.   64 pages of the specification, including 71 claims and an abstract.

2.   11 sheets of drawings.

3.   79 pages of Sequence Listing.

4.   An oath or declaration of the inventors (unsigned).

The filing fee is being deferred at this time.

Dated: January 5, 2000.

[signature]

Tom Hunter (Reg. No. 38,498)
MAJESTIC, PARSONS, SIEBERT & HSUE P.C.
Four Embarcadero Center, Suite 1100
San Francisco, California 94111-4106
Telephone: (415) 248-5500
Facsimile: (415) 362-5418

Atty. Docket: 2500.128US1
UC Ref: 99-174-1

-2-

In the United States Patent and Trademark Office

U.S. Patent Application For

GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR ANTIBIOTIC C-1027

Inventor(s):   **BEN SHEN**, a citizen of the Peoples Republic of China, residing at 1842 Rushmore Lane, Davis, CA 95616, USA

WEN LIU, a citizen of the Peoples Republic of China, residing at the Institute of Medicinal biotechnology, Tiantan, Beijing, 100005, China

STEVEN D. CHRISTENSON, a citizen of the United States of America, residing at 1079 Monarch Lane, Davis, CA, 95616, USA

SCOTT STANDAGE, a citizen of the United Kingdom, residing at 63 Tudor Road, Bornet, Herts, EN5 5NW, U.K.

Assignee: The Regents of the University of California

Entity: Small Entity

MAJESTIC, PARSONS, SIEBERT & HSUE P.C.
Four Embarcadero Center, Suite 1100
San Francisco, CA 94111-4106
Tel: 415 248-5500
Fax: 415 362-5418

GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR ANTIBIOTIC C-1027

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit under 35 U.S.C. §119 of provisional
5   application USSN 60/115,434, filed on January 6, 1999, which is herein incorporated by
reference in its entirety for all purposes.

FIELD OF THE INVENTION

This invention relates to the field of enediyne antibiotics. In particular this
15   invention elucidates the gene cluster controlling the biosynthesis of the C-1027 enediyne.

BACKGROUND OF THE INVENTION

The enediyne antibiotics are currently the focus of intense research activity in
the fields of chemistry, biology, and medical sciences, because of their unique molecular
architecture, biological activities, and modes of actions (Doyle and Borders (1995) *Enediyne*
20   *antibiotics as antitumor agents*. Marcel-Dekker, New York, Thorson *et al.* (1999) *Bioorg.*
Chem., 27: 172-188). Since the unveiling of the structure of neocarzinostatin chromophore
(Edo *et al.* (1985) *Tetrahedron Lett.* 26: 331-340) in 1985, the enediyne family has grown
steadily. Thus far, there have been three basic groups within the enediyne antibiotic family:
(a) the calicheamicin/esperamicin type, which includes the calicheamicins, the esperamicins,
25   and namenamicin, (b) the dynemicin type, and (c) the chromoprotein type, consisting of an
apoprotein and an unstable enediyne chromophore. The latter group includes
neocarzinostatin, kedarcidin, C-1027 (Fig. 1), and maduropeptin, whose enediyne
chromophore structures have been established, as well as several others whose enediyne
chromophore structures are yet to be determined due to their instability (Thorson *et al.*

(1999) *Bioorg. Chem.*, 27: 172-188). N1999A2, in contrast to the other chromoproteins, exists as an enediyne chromophore alone despite the fact that its structure is very similar to the other chromoprotein chromophore (*Ando et al.*(1998) *Tetra. Letts.,* 39: 6495-6480).

As a family, the enediyne antibiotics are the most potent, highly active
5   antitumor agents ever discovered. Some members are 1000 times more potent than adriamycin, one of the most effective, clinically used antitumor antibiotics (Zhen *et al.* (1989) *J. Antibiot.* 42: 1294-1298). All members of this family contain a unit consisting of two acetylenic groups conjugated to a double bond or incipient double bond within a nine or ten-membered ring; i.e., the enediyne core as exemplified by C-1027 in Fig. 1. As the
10   consequence of this structural feature, these compounds share a common mechanism of action: the enediyne core undergoes an electronic rearrangement to form a transient benzenoid diradical, which is positioned in the minor groove of DNA so as to damage DNA by abstracting hydrogen atoms from deoxyriboses on both strands (Fig. 1). Reaction of the resulting deoxyribose carbon-centered radicals with molecular oxygen initiates a process that
15   results in both single-strand and double-strand DNA cleavages (Doyle and Borders (1995) *Enediyne antibiotics as antitumor agents.* Marcel-Dekker, New York; Ikemoton *et al.* (1995) *Proc. Natl. Acad. Sci. USA* 92:10506-10510; Myers *et al.* (1997) *J. Am. Chem. Soc.* 119: 2965-2972; Stassinopoulos *et al.* (1996) *Science* 272: 1943-1946; Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188; Xu *et al.* (1997) *J. Am. Chem. Soc.* 119: 1133-1134). This
20   novel mechanism of DNA damage has important implications for their application as potent cancer chemotherapeutic agents (Doyle and Borders (1995) *supra.*; Sievers *et al.* (1999) *Blood* 93: 3678-3684).

As an alternative to making structural analogs of microbial metabolites by chemical synthesis, manipulations of genes governing secondary metabolism offer a
25   promising alternative allowing preparation of these compounds biosynthetically (Cane *et al.* (1998) *Science* 282: 63-68; Hutchinson and Fujii. (1995) *Ann. Rev. Microbiol.* 49: 201-38; Katz and Donadio (1993) *Ann. Rev. Microbiol.* 47: 875-912). The success of the latter approach depends critically on the availability of novel genetic systems and on genes encoding novel enzyme activities. The enediynes offer a distinct opportunity to study the
30   biosynthesis of their unique molecular scaffolds and the mechanism of self-resistance to extremely cytotoxic natural products. Elucidation of these aspects provides access to rational engineering of enediyne biosynthesis for novel drug leads and makes it possible to construct enediyne overproducing strains by de-regulating the biosynthetic machinery. In

-2-

addition, elucidation of an enediyne gene cluster contributes to the general field of combinatorial biosynthesis by expanding the repertoire of novel polyketide synthase (PKS) and deoxysugar biosynthesis genes as well as other genes uniquely associated with enediyne biosynthesis, leading to the making of novel enediynes via combinatorial biosynthesis.

SUMMARY OF THE INVENTION

This invention provides nucleic acid sequences and characterization of the gene cluster responsible for the biosynthesis of the enediyne C-1027 (produced by *Streptomyces globisporus*). In particular structural and functional characterization is provided for the 50 open reading frames (ORFs) comprising this gene cluster. Thus, in one embodiment, this invention provides an isolated nucleic acid comprising a nucleic acid selected from the group consisting of a nucleic acid encoding any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA), a nucleic acid encoding a polypeptide encoded by any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA); and a nucleic acid amplified by polymerase chain reaction (PCR) using primer pairs that amplify any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA). In one embodiment, preferred nucleic acids comprise a nucleic acid encoding at least two (more preferably at least three or more) open reading frames (ORFs) selected from the group consisting of ORF-1 through ORF 42, excluding ORF 9 (cagA).

In another embodiment this invention provides an isolated nucleic acid comprising a nucleic acid that specifically hybridizes under stringent conditions to an open reading frame (ORF) of the C-1027 biosynthesis gene cluster, excluding ORF 9 (cagA), and can substitute for the ORF to which it specifically hybridizes to direct the synthesis of an enediyne. In certain embodiments this also includes nucleic acids that would stringently hybridizes indicated above, but for, the degeneracy of the nucleic acid code. In other words, if silent mutations could be made in the subject sequence so that it hybridizes to he indicated sequence(s) under stringent conditions, it would be included in certain embodiments. Particularly preferred nucleic acids comprises a nucleic acid that specifically hybridizes under stringent conditions to a nucleic acid selected from the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 8, ORF 10, ORF 11, ORF 12, ORF 13, ORF 14, ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF

-3-

26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42. Particularly preferred isolated nucleic acid comprises a nucleic acid selected from the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4,

5   ORF 5, ORF 6, ORF 7, ORF 8, ORF 10, ORF 11, ORF 12, ORF 13, ORF 14, ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42. The nucleic acid may comprises a nucleic acid that is a single nucleotide polymorphism (SNP) of a nucleic acid

10   selected from the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 81, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 8, ORF 10, ORF 11, ORF 12, ORF 13, ORF 14, ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33,

15   ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42.

This invention also provides an isolated gene cluster comprising open reading frames encoding polypeptides sufficient to direct the assembly of a C-1027 enediyne or a C-1027 enediyne analogue. The gene cluster may be present in a cell, more preferably in a bacterial cell (*e.g. Actinomycetes, Actinoplanetes, Actinomadura, Micromonospora,* or

20   *Streptomycetes*). Particular preferred bacterial cells include, but are not limited to *Streptomyces globisporus, Streptomyces lividans, Streptomyces coelicolor, Micromonospora echinospora spp. calichenisis, Actinomadura verrucosopora, Micromonospora chersina, Streptomyces carzinostaticus,* and *Actinomycete* L585-6. The gene cluster may contain one or more open reading frames is operatively linked to a heterologous promoter (*e.g.* a

25   constitutive or an inducible promoter).

This invention also provides for an polypeptide encoded by any one or more of the nucleic acids described herein.

Also provided are host cell(s) (*e.g.* eukaryotic cells or bacterial cells as described herein) transformed with one or more of the expression vectors described herein.

30   Preferred host cells are transformed with an exogenous nucleic acid comprising a gene cluster encoding polypeptides sufficient to direct the assembly of a C-1027 enediyne or a C-1027 enediyne analogue. In certain embodiments, heterologous nucleic acid may comprise only a portion of the gene cluster, but the cell will still be able to express an enediyne.

-4-

This invention also provides methods of chemically modifying a biological molecule. The methods involve contacting a biological molecule that is a substrate for a polypeptide encoded by a C-1027 biosynthesis gene cluster open reading frame, with a polypeptide encoded by a C-1027 biosynthesis gene cluster open reading frame whereby the

5   polypeptide chemically modifies the biological molecule. In one preferred embodiment, the polypeptide is an enzyme selected from the group consisting of a hydroxylase, a homocysteine synthase, a dNDP-glucose dehydrogenase, a citrate carrier protein, a C-methyl transferase, an N-methyl transferase, an aminotransferase, a CagA apoprotein, an NDP-glucose synthase, an epimerase, an acyl transferase, a coenzyme F390 synthase, and

10   epoxidase hydrolase, an anthranilate synthase, a glycosyl transferase, a monooxygenase, a type II condensation protein, an aminomutase, a type II adenylation protein, an O-methyl transferase, a P-450 hydroxylase, an oxidoreductase, and a proline oxidase. In a preferred embodiment the method involves contacting the biological molecule with at least two (preferably at least three or more) different polypeptides encoded by C-1027 biosynthesis

15   gene cluster open reading frames. The contacting may be in a host cell (e.g. a eukaryotic cell or a bacterial cell) or the contacting can be ex vivo. The biological molecule can be an endogenous metabolite produced by said host cell or an exogenous supplied metabolite. In preferred embodiments, the host cell is a bacterial cell or eukaryotic cell (e.g., a mammalian cell, a yeast cell, a plant cell, a fungal cell, an insect cell, etc.). In certain preferred

20   embodiments, the host cell synthesizes sugars and glycosylates the biological molecule. In other preferred embodiments, the host cell synthesizes deoxysugars. The method can further involve contacting the biological molecule with a polyketide synthase or a non-ribosomal polypeptide synthetase. The contacting can be in a cell (e.g., a bacterial cell) or ex vivo. In one preferred embodiment the method comprises contacting the biological molecule with at

25   substantially all of the polypeptides encoded by C-1027 biosynthesis gene cluster open reading frames and said method produces an enediyne or enediyne analogue. In another preferred embodiment, the biological molecule is a fatty acid and the biological molecule is contacted with a C-1027 orf polyeptide selected from the group consisting of an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an

30   oxidoreductase, and a proline oxidase. In certain embodiments, the biological molecule is a fatty acid and said biological molecule is contacted with a plurality of C-1027 orf polypeptides comprising an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase. In one especially preferred

embodiment ,the biological molecule is contacted with polypeptides encoded by ORF17, ORF20, ORF21, ORF29, ORF30, ORF32, ORF35, and ORF38. In another especially preferred embodiment, the biological molecule is contacted with polypeptides encoded by ORF 15, ORF 16, ORF 28, ORF3, ORF 14, and ORF 13, and, in certain embodiments, ORF

5   4 and ORF 3 as well.

In certain embodiments, the method may comprise contacting a sugar with one or more C-1027 open reading frame polypeptides selected from the group consisting of a dNDP-glucose synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase. Particularly

10   preferred variant of this method comprise contacting a dNDP-glucose with a plurality of C-1027 open reading frame polypeptides comprising a dNDP-glucose synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase.

In certain other embodiments, the method comprises contacting an amino acid

15   with one or one or more C-1027 open reading frame polypeptides selected from the group consisting of a hydroxylase, an aminomutase, a type II NRPS condensation enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein. These methods may involve contacting an amino acid with a plurality of C-1027 open reading frame polypeptides comprising a hydroxylase, a halogenase, an aminomutase, a type II NRPS condensation

20   enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein. In particularly preferred embodiments, the amino acid is a tyrosine.

This invention also provides a method of synthesizing a chromaprotein type enediyne core, said method comprising contacting a fatty acid with one or more C-1027 orf polypeptides selected from the group consisting of an epoxide hydrase, a monooxygenase, an

25   iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase. In preferred embodiments, the fatty acid may be contacted with a plurality of C-1027 orf polypeptides comprising an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase. In particularly preferred embodiments, the fatty acid is contacted with polypeptides encoded by ORF17, ORF20,

30   ORF21, ORF29, ORF30, ORF32, ORF35, and ORF38.

In still yet another embodiment, this invention provides a method of synthesizing a deoxysugar. This method involves contacting a sugar with one or more C-1027 open reading frame polypeptides selected from the group consisting of a dNDP-glucose

-6-

synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase. In preferred embodiments, this method involves contacting a dNDP-glucose with a plurality of C-1027 open reading frame polypeptides comprising a dNDP-glucose synthase, a dNDP glucose

5   dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase. In particularly preferred embodiments, the dNDP-glucose is contacted with polypeptides encoded by ORF17, ORF20, ORF21, ORF29, ORF30, ORF32, ORF35, and ORF38.

This invention also provides methods of synthesizing a beta amino acid by

10 contacting an amino acid with one or one or more C-1027 open reading frame polypeptides selected from the group consisting of a hydroxylase, an aminomutase, a type II NRPS condensation enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein. The method preferably comprises contacting an amino acid with a plurality of C-1027 open reading frame polypeptides comprising a hydroxylase, a halogenase, an

15 aminomutase, a type II NRPS condensation enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein. Particularly preferred embodiments comprise contacting the amino acid (e.g. tyrosine) with polypeptides encoded by ORF 4, ORF11, ORF24, ORF23, ORF25, and ORF26.

Also provided are methods of synthesizing an enediyne or an enediyne

20 analogue. These methods involve culturing a cell (e.g. a eukaryotic cell or a bacterium) comprising a recombinantly modified C-1027 gene cluster under conditions whereby said cell expresses said enediyne or enediyne analogue; and recovering the enediyne or enediyne analogue. In preferred embodiments, the gene cluster is present in a bacterium (e.g., Actinomycetes, Actinoplanetes, Actinomadura, Micromonospora, or Streptomycetes).

25 Particularly preferred bacteria include, but are not limited to Streptomyces globisporus, Streptomyces lividans, Streptomyces coelicolor, Micromonospora echinospora spp. calichenisis, Actinomadura verrucosopora, Micromonospora chersina, Streptomyces carzinostaticus, and Actinomycete L585-6. In another preferred embodiment, the gene cluster is present in a eukaryotic cell (e.g. a mammalian cell, a yeast cell, a plant cell, a

30 fungal cell, an insect cell, etc.). The host cell can be one that synthesizes sugars and glycosylates the enediyne or enediyne analogue. The host can be one that synthesizes deoxysugars.

This invention also provides a method of making a cell (*e.g.,* a bacterial or eukaryotic cell) resistant to an enediyne or an enediyne metabolite. This method involves expressing in the cell one or more isolated C-1027 open reading frame nucleic acids that encode a protein selected from the group consisting of a CagA apoprotein, a SgcB

5   transmembrane efflux protein, a transmembrane transport protein, a Na+/H+ transporter, an ABC transport, a glycerol phosphate tranporter, and a UvrA-like protein. In preferred embodiments, the isolated C-1027 open reading frame nucleic acids are selected from the group consisting of ORF 9, ORF2, ORF 27, ORF 0, ORF 1 c-terminus, ORF 2, and ORF 1 N-terminus. Certain embodiments exclude cagA (ORF 9).

10   In one embodiment, this invention specifically excludes one or more of open reading frames -7 through 42. In particular, in one embodiment this invention excludes cagA (ORF 9), and/or sgcA (ORF 1), and/or sgcB (ORF 2).

DEFINITIONS

The terms "C-1027 open reading frame", and "C-1027 ORF" refer to an open

15   reading frame in the C-1027 biosynthesis gene cluster as isolated from *Streptomyces globisporus.* The term also embraces the same open reading frames as present in other enediyne-synthesizing organisms (*e.g.* other strains and/or species of *Streptomyces, Actinomyces,* and the like). The term encompasses allelic variants and single nucleotide polymorphisms (SNPs). In certain instances the C-1027 ORF is used synonymously with the

20   polypeptide encoded by the C-1027 ORF and may include conservative substitutions in that polypeptide. The particular usage will be clear from context.

The terms "isolated" "purified" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. With respect to nucleic acids and/or polypeptides the term can refer to

25   nucleic acids or polypeptides that are no longer flanked by the sequences typically flanking them in nature.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a

30   corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The term also includes variants on the traditional peptide linkage joining the amino acids making up the polypeptide.

The terms "nucleic acid" or "oligonucleotide" or grammatical equivalents herein refer to at least two nucleotides covalently linked together. A nucleic acid of the present invention is preferably single-stranded or double stranded and will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are

5   included that may have alternate backbones, comprising, for example, phosphoramide (Beaucage et al. (1993) Tetrahedron 49:1925) and references therein; Letsinger (1970) J. Org. Chem. 35:3800; Sprinzl et al. (1977) Eur. J. Biochem. 81: 579; Letsinger et al. (1986) Nucl. Acids Res. 14: 3487; Sawai et al. (1984) Chem. Lett. 805, Letsinger et al. (1988) J. Am. Chem. Soc. 110: 4470; and Pauwels et al. (1986) Chemica Scripta 26: 141 9),

10   phosphorothioate (Mag et al. (1991) Nucleic Acids Res. 19:1437; and U.S. Patent No. 5,644,048), phosphorodithioate (Briu et al. (1989) J. Am. Chem. Soc. 111 :2321, O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm (1992) J. Am. Chem. Soc. 114:1895; Meier et al. (1992) Chem. Int. Ed.

15   Engl. 31: 1008; Nielsen (1993) Nature, 365: 566; Carlsson et al. (1996) Nature 380: 207). Other analog nucleic acids include those with positive backbones (Denpcy et al. (1995) Proc. Natl. Acad. Sci. USA 92: 6097; non-ionic backbones (U.S. Patent Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Angew. (1991) Chem. Intl. Ed. English 30: 423; Letsinger et al. (1988) J. Am. Chem. Soc. 110:4470; Letsinger et al. (1994) Nucleoside

20   & Nucleotide 13:1597; Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook; Mesmaeker et al. (1994), Bioorganic & Medicinal Chem. Lett. 4: 395; Jeffs et al. (1994) J. Biomolecular NMR 34:17; Tetrahedron Lett. 37:743 (1996) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series

25   580, Carbohydrate Modifications in Antisense Research, Ed. Y.S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al. (1995), Chem. Soc. Rev. pp169-176). Several nucleic acid analogs are described in Rawls, C & E News June 2, 1997 page 35. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of

30   additional moieties such as labels, or to increase the stability and half-life of such molecules in physiological environments.

The term "heterologous" as it relates to nucleic acid sequences such as coding sequences and control sequences, denotes sequences that are not normally associated with a

region of a recombinant construct, and/or are not normally associated with a particular cell. Thus, a "heterologous" region of a nucleic acid construct is an identifiable segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a

5   construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Similarly, a host cell transformed with a construct which is not normally present in the host cell would be considered heterologous for

10   purposes of this invention.

A "coding sequence" or a sequence which "encodes" a particular polypeptide (*e.g.* a PKS, an NRPS, *etc.*), is a nucleic acid sequence which is ultimately transcribed and/or translated into that polypeptide *in vitro* and/or *in vivo* when placed under the control of appropriate regulatory sequences. In certain embodiments, the boundaries of the coding

15   sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from procaryotic or eucaryotic mRNA, genomic DNA sequences from procaryotic or eucaryotic DNA, and even synthetic DNA sequences. In preferred embodiments, a transcription termination sequence will usually be located 3' to the coding sequence.

20   Expression "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell. Not all of these control sequences need always be present in a recombinant vector so long as the desired gene is

25   capable of being transcribed and translated.

"Recombination" refers to the reassortment of sections of DNA or RNA sequences between two DNA or RNA molecules. "Homologous recombination" occurs between two DNA molecules which hybridize by virtue of homologous or complementary nucleotide sequences present in each DNA molecule.

30   The terms "stringent conditions" or "hybridization under stringent conditions" refers to conditions under which a probe will hybridize preferentially to its target subsequence, and to a lesser extent to, or not at all to, other sequences. "Stringent hybridization" and "stringent hybridization wash conditions" in the context of nucleic acid

-10-

hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes part I chapter*

5 *2 Overview of principles of hybridization and the strategy of nucleic acid probe assays*, Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very

10 stringent conditions are selected to be equal to the T_m for a particular probe.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is

15 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see*, Sambrook *et al.* (1989) *Molecular Cloning - A Laboratory Manual (2nd ed.)* Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency

20 wash for a duplex of, *e.g.*, more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids which do not hybridize to each other under stringent conditions

25 are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Expression vectors are defined herein as nucleic acid sequences that are direct the transcription of cloned copies of genes/cDNAs and/or the translation of their mRNAs in

30 an appropriate host. Such vectors can be used to express genes or cDNAs in a variety of hosts such as bacteria, bluegreen algae, plant cells, insect cells and animal cells. Expression vectors include, but are not limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses. Specifically designed vectors allow the shuttling of DNA

-11-

between hosts, such as bacteria-yeast or bacteria-animal cells. An appropriately constructed expression vector preferably contains: an origin of replication for autonomous replication in a host cell, a selectable marker, optionally one or more restriction enzyme sites, optionally one or more constitutive or inducible promoters. In preferred embodiments, an expression vector is a replicable DNA construct in which a DNA sequence encoding a one or more PKS and/or NRPS domains and/or modules is operably linked to suitable control sequences capable of effecting the expression of the products of these synthase and/or synthetases in a suitable host. Control sequences include a transcriptional promoter, an optional operator sequence to control transcription and sequences which control the termination of transcription and translation, and so forth.

The term "conservative substitution" is used in reference to proteins or peptides to reflect amino acid substitutions that do not substantially alter the activity (specificity or binding affinity) of the molecule. Typically conservative amino acid substitutions involve substitution one amino acid for another amino acid with similar chemical properties (e.g. charge or hydrophobicity). The following six groups each contain amino acids that are typical conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

The "group consisting of ORF-1 through ORF 42" refers to the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 81, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 8, ORF 9, ORF 10, ORF 11, ORF 12, ORF 13, ORF 14, ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42 as identified in Tables I and II. In certain embodiments ORF 9 (cagA) is excluded.

A "biological molecule that is a substrate for a polypeptide encoded by a enediyne (*e.g.*, C-1027) biosynthesis gene" refers to a molecule that is chemically modified by one or more polypeptides encoded by open reading frame(s) of the C-1027 biosynthesis gene cluster. The "substrate" may be a native molecule that typically participates in the biosynthesis of an enediyne, or can be any other molecule that can be similarly acted upon by the polypeptide.

-12-

A "polymorphism" is a variation in the DNA sequence of some members of a species. A polymorphism is thus said to be "allelic," in that, due to the existence of the polymorphism, some members of a species may have the unmutated sequence (*i.e.* the original "allele") whereas other members may have a mutated sequence (*i.e.* the variant or mutant "allele"). In the simplest case, only one mutated sequence may exist, and the polymorphism is said to be diallelic. In the case of diallelic diploid organisms, three genotypes are possible. They can be homozygous for one allele, homozygous for the other allele or heterozygous. In the case of diallelic haploid organisms, they can have one allele or the other, thus only two genotypes are possible. The occurrence of alternative mutations can give rise to trialleleic, *etc.* polymorphisms. An allele may be referred to by the nucleotide(s) that comprise the mutation.

"Single nucleotide polymorphism" or "SNPs are defined by their characteristic attributes. A central attribute of such a polymorphism is that it contains a polymorphic site, "X," most preferably occupied by a single nucleotide, which is the site of the polymorphism's variation (Goelet and Knapp U.S. patent application Ser. No. 08/145,145). Methods of identifying SNPs are well known to those of skill in the art (*see, e.g.,* U.S. Patent 5,952,174).

Abbreviations used herein include LB, Luria-Bertani; NGDH, dNDP-glucose 4,6-dehydratase ; nt, nucleotide; ORF, open reading frame; PCR, polymerase chain reaction; PEG, polyethyleneglycol; PKS, polyketide synthase; RBS, ribosomal binding site; Apr, apramycin; R, resistant; Th, thiostrepton; WT, wild-type; and TS, temperature sensitive

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the structures of C-1027 chromophore and the benzenoid diradical intermediate proposed to initiate DNA cleavage.

Figure 2 illustrates a scheme using C-1027 open reading frame polypeptides for the synthesis of deoxysugars.

Figure 3A illustrates a scheme using C-1027 open reading frame polypeptides for the synthesis of a β-amino acid.

Figure 3B illustrates a scheme using C-1027 open reading frame polypeptides for the synthesis of a benzoxazolinate.

Figure 4 illustrates the synthesis of the enediyne core and final assembly of the C-1027 enediyne.

Figures 5A, 5B, and 5C illustrate the organization of the C-1027 enediyne biosynthetic gene cluster. Figure 5A shows a restriction map of the 75-kb *sgc* gene cluster from *S. globisporus* as represented by three cosmid clones. Figure 5B illustrates the genetic organization of the *sgcA, sgcB*, and *cagA* genes, showing that they are clustered in the *sgc* gene cluster. Probe 1, the 0.55-kb dNDP-glucose 4,6-dehydratase gene fragment from pBS1002. Probe 2, the 0.73-kb *cagA* fragment from pBS1003. A, *Apa*I; B, *Bam*HI; E, *Eco*RI; K, *Kpn*I, S, *Sac*II; Sp, *Sph*I. Figure 5C shows the genetic organization of the C-1027 biosynthesis gene cluster.

Figure 6 shows the DNA and deduced amino acid sequences of the 3.0-kb *Bam*HI fragment from pBS1007, showing the *sgcA* and *sgcB* genes. Possible RBSs are boxed. The presumed translational start and stop sites are in boldface. Restriction enzyme sites of interest are underlined. The amino acids, according to which the degenerated PCR primer were designed for amplifying the dNDP-glucose 4,6-dehydratase gene from *S. globisporus*, are underlined.

Figure 7 shows the amino acid sequence alignment of SgcA with three other dNDP-glucose 4,6-dehydratases. Gdh, TDP-glucose 4,6-dehydratase of *S. erythraea* (AAA68211); MtmE, TDP-glucose 4,6-dehydratase in the mithramycin pathway of *S. argillaceus* (CAA71847); TylA2, TDP-glucose 4,6-dehydratase in the tylosin pathway of *S. fradiae* (S49054). Given in parentheses are protein accession numbers. The αβα fold with the NAD$^+$-binding motif of GxGxxG is boxed.

Figures 8A and 8B show disruption of *sgcA* by single crossover homologous recombination. Figure 8A shows construction of *sgcA* disruption mutant and restriction maps of the wild-type *S. globisporus* C-1027 and *S. globisporus* SB1001 mutant strains showing predicted fragment sizes upon *Bam*HI digestion. Figures 8B and 8C show a Southern analysis of *S. globisporus* C-1027 (lane 1) and *S. globisporus* SB1001 (lanes 2, 3, and 4, three individual isolates) genomic DNA, digested with *Bam*HI, using (Figure 8B) pOJ260 vector or (Figure 8C) the 0.75-kb *Sac*II/*Kpn*I fragment of *sgcA* from pBS1012 as a probe, respectively. B, *Bam*HI; K, *Kpn*I; S, *Sac*II.

Figures 9A, 9B, 9B, and 9D illustrate the determination of C-1027 production in various *S. globisporus* strains by assaying their antibacterial activity against *M. luteus*.

-14-

Figure 9A:1, *S. globisporus*C-1027; 2,3, and 4, *S. globisporus* SB1001 (three individual isolates); 5, *S. globisporus* AF67; 6, *S. globisporus* AF40. Figure 9B: 1, *S. globisporus* C-1027; 2, *S. globisporus* SB1001 (pWHM3); 3 and 4, *S. globisporus* SB1001 (pBS1015) (two individual isolates). Both *S. globisporus* SB1001 (pWHM3) and *S. globisporus* SB1001

5   (pBS1015) were grown in the presence of 5 µg/mL thiostrepton. Figure 9C: 1, *S. globisporus*C-1027; 2, *S. globisporus* SB1001 (pBS1015); 3. *S. globisporus* SB1001; 4. *S. globisporus* SB1001 (pWHM3); 5. *S. globisporus* AF40; 6. *S. globisporus* AF44. All *S. globisporus* strains were grown in the absence of thiostrepton. Figure 9D: 1. *S. globisporus* (pKC1139); 2. *S. globisporus* (pBS1018).

10   **DETAILED DESCRIPTION**

This invention provides a complete gene cluster regulating the biosynthesis of C-1027, the most potent member of the enediyne antitumor antibiotic family. C-1027 is produced by *Streptomyces globisporus* C-1027 and consists of an apoprotein (encoded by the *cagA* gene) and a non-peptidic chromophore. The C-1027 chromophore could be viewed as

15   being derived biosynthetically from a benzoxazolinate, a deoxyamino hexose, a β-amino acid, and an enediyne core. Adopting a strategy to clone the C-1027 biosynthesis gene cluster by mapping a putative dNDP-glucose 4,6-dehydratase (NGDH) gene to *cagA*, we localized 75 kb contiguous DNA from *S. globisporus* encoding a complete C-1027 gene cluster.

20   Initial sequencing of the cloned gene cluster revealed two genes, *sgcA* and *sgcB*, that encode an NGDH enzyme and a transmembrane efflux protein, respectively, and confirmed that the *cagA* gene resides approximately 14 kb upstream of the *sgcA,B* locus. The involvement of the cloned gene cluster in C-1027 biosynthesis was demonstrated by disrupting the *sgcA* gene to generate C-1027-nonproducing mutants and by complementing

25   the *sgcA* mutants in vivo to restore C-1027 production.

Subsequent DNA sequence analysis provided the complete enediyne C-1027 gene cluster sequence (SEQ ID NOs: 1 and 2) revealing 50 open reading frames which are summarized in Tables I and II. These results represent the first cloning of a gene cluster for enediyne anti-tumor antibiotic biosynthesis.

30

Table I. Summary of the C-1027 gene cluster open reading frames. Table 1. C-1027 gene cluster open reading frames (-7 to 26), primers for ORF amplification, and proposed functions

orf #	Size	Relative position	Primers		Function	Seq ID No.
orf-(-7)	648 bp	658-11	Fwd: ATG GGC ATG ACG GGT	Rev: CTA GAG GAT CCC GGG	very weak homology to putative hydroxylase	3 4
orf-(-6)	549 bp	1478-930	Fwd: ATG CCG CGG ATT CCC	Rev: TCA GCT GTC GAT GTC	Viral infectivity potentiator protein	5 6
orf-(-5)	1065 bp	2713-1649	Fwd: ATG ACC ATC GCC ACT	Rev: TCA GAG GCC GAG CAC	N-truncated Methionine synthase (likely psuedogene)	7 8
orf-(-4)	387 bp	3238-2851	Fwd: ATG AGC TCG CTA CTG	Rev: CTA GGA GCC GGT CGC	Viral transcription factor	9 10
orf-(-3)	1530 bp	4971-3442	Fwd: ATG AGC AGC AGC GCC	Rev: TCA TTC GTC GGC TGC	Viral Homolog possibly primase	11 12
orf-(-2)	3027 bp	5982-7478	Fwd: GTG AGG GCT CTG CCG	Rev: TCA GAC GGC GGA GGG	Glycerol-Phosphate ABC Transporter (SnoX drug resistance)	13 14
orf-(-1)	2328 bp	9900-7573	Fwd: GTG AGC GTC ACC GAC	Rev: TCA ACC CGC CCT GCG	UvrA-like drug resistance pump	15 16
orf-0	1368 bp	11349-9982	Fwd: ATG AGG ATG CTG GTG	Rev: GTG GCT GTG CTC GCA	Na$^+$/H$^+$ efflux pump	17 18
orf-1	999 bp	28590-29588	Fwd: ATG AGG ATG CTG GTG	Rev: TCA GCC GAC GGC GTC	dNTP-glucose dehydratase	19 20
orf-2	1566 bp	29632-31197	Fwd: GTG ACA GCA GTC AAG	Rev: TCA TGT GGC CGG TTG	Transmembrane efflux protein	21 22
orf-3	1311 bp	31280-32590	Fwd: GTG GAG TAC TGG AAC	Rev: TCA GGC CTG AGG GGC	Coenzyme F390 synthase phenylacetyl-CoA ligase	23 24
orf-4	1584 bp	32809-34392	Fwd: GTG CCC CAC GGT GCA	Rev: CTA CAG CCC TCC GAG	phenol hydroxylase chlorophenol-4-monooxygenase	25 26
orf-5	bp	35274-34458	Fwd: ATG TCT TCA ACC CGT	Rev: TCA GCC GCG CAG GAA	citrate transport protein	27 28
orf-6	1272	17924-16653	Fwd: ATG CTG GAG AAA TGC	Rev: TCA GAC GAG CTC CTT	C-methyl transferase hydroxylase	29 30
orf-	735	16653-	Fwd: ATG GAG TAC GGC CCC		N-	31

7	bp	15919	Rev: TCA TGC CGT GCG CAC	methyltransferas e	32
orf-8	1233 bp	15922-14690	Fwd: ATG AGC GGC GGC CCG Rev: TCA CCT CGC CGG ACG	Aminotransferase	33 34
orf-9	432 bp	14643-14212	Fwd: ATG TCG TTA CGT CAC Rev: TCA GCC GAA GGT CAG	CagA	35 36
orf-10	1068 bp	13012-14079	Fwd: ATG AAG GCA CTT GTA Rev: TCA GGC CGC GAT CTC	dNTP-glucose synthase	37 38
orf-11	1485 bp	12835-11351	Fwd: GTG GAC GTG TCA GCG Rev: TCA GGA CCG CGC ACC	Hydroxylase, Halogenase	39 40
orf-12	579 bp	25564-24986	Fwd: ATG AAG CCG ATC GGG Rev: TCAGGA CGA CTT GTT	dNTP-4-keto-6-deoxyglucose 3,5-epimerase	41 42
orf-13	1137 bp	24702-23566	Fwd: ATG CCT TCC CCC TTC Rev: TCA GGT GCG CTC GGC	3-O-acyltransferase	43 44
orf-14	1455 bp	22878-21424	Fwd: GTG AGA GAC GGC CGG Rev:TCA CGT GGT GAT GGC	Coenzyme F-390 Synthase Phenylacetyl CoA Ligase	45 46
orf-15	1482 bp	21407-19926	Fwd: ATG ACC GAC CAG TGC Rev: TCA CAG CAA CTC CTC	Anthranilate Synthase I	47 48
orf-16	663 bp	19929-19267	Fwd: GTG AGC TTG TGG TCT Rev: TCA GGC CGG TTC GGC	Anthranilate Synthase II	49 50
orf-17	1161 bp	19191-18031	Fwd: GTG CGT CCC TTC CGT Rev: TCA GCG GAG CGG ACG	epoxide hydrolase	51 52
orf-18	423 bp	35938-35516	Fwd: ATG CCA GCA CCG ACT Rev: TCA GTC GTT GCC GCG	Unknown	53 54
orf-19	1380 bp	27214-28593	Fwd: ATG CGG GTG ATG ATC Rev: TCA TCG GTC CGC CTC	glycosyl transferase	55 56
orf-20	1356 bp	25815-27170	Fwd: ATG ACC AAG CAC GCC Rev: TCA TAC GGC GGC GCC	squalene monooxygenase	57 58
orf-21	672 bp	23546-22875	Fwd: GTG AGC GCA CAA CTC Rev: TCA CGG CTG TGC CTG	hypothetical Fe-S flavoprotein	59 60
orf-22	816 bp	35274-34458	Fwd: ATG TCT TCA ACC CGT Rev: TCA GCC GCG CAG GAA	haloacetate dehalogenase hydrolase	61 62
orf-23	1380 bp	37559-38938	Fwd: ATG ACG ACG TCC GAC Rev: TCA GGA GGT GAA GGG	peptide synthetase	63 64
orf-24	1620 bp	40986-39367	Fwd: ATG GCA TTG ACT CAA Rev: TCA GCG CAG CTG GAT	Histidine Ammonia lyase	65 66
orf-25	1560 bp	42611-41052	Fwd: ATG ACG CGG CCG GTG Rev: TCA GCG GGT GAG CCG	Type II adenylation protein	67 68
orf-26	282 bp	38983-39264	Fwd: GTG TCC ACC GTT TCC Rev: TCA CTG CGT TCC GGA	Type II peptidyl carrier protein	69 70

Table II. C-1027 gene cluster open reading frames (27 to 42), primers for ORF amplification, and proposed functions

ORF	Relative Position	Primers		Function	SEQ ID NO.
orf-27	43945-46023	Fwd:	GTG TGC CCG GTG ACA GAC	Antibiotic	71
		Rev:	TCA GCC CAC GGG CTG GGA	Transporter	72
orf-28	46167-47171	Fwd:	GTG TTG GGC GAT GAG GAC	O-	73
		Rev:	TCA GAC CGC GGA CAT CTG	methyltransfer ase	74
orf--29	47227-48485	Fwd:	ATG GCC GGC CTG GTC ATG	p450	75
		Rev:	TCA GGA CCC GAG GGT CAC	hydroxylase	76
orf-30	48610-49714	Fwd:	GTG GAC CAG ACG TCT ACG	Oxidoreductase	77
		Rev:	TCA TGC AGG TGC AGC GTG		78
orf-31	50350-51390	Fwd:	ATG AGG CCG CTC GTT CGG	Unknown	79
		Rev:	TCA TCC CGG CCC GGC GGC	Protein	80
orf-32	51420-52341	Fwd:	ATG AGA ACG CGG CGA CGC	Oxidoreductase	81
		Rev:	TCA CGG CCG GAG GCG TAC		82
orf-33	53241-54074	Fwd:	GTG TAT CAG CCG GAC TGT	Unknown	83
		Rev:	CTA CTC ATT CCA GTT GTG	Protein	84
orf-34	54230-55379	Fwd:	ATG TCT ACG GGC TAT CTC	Unknown	85
		Rev:	TCA GCC GCC GGT GGC GCC	Protein	86
orf-35	56027-56881	Fwd:	ATG TTC TCC CCC GCC GCC	Oxidase/	87
		Rev:	TCA GTA CGC CTG GTG GGC	Dehydrogenase	88
orf-36	56928-57730	Fwd:	ATG AAT TCG CTC GAC GAC	Unknown	89
		Rev :	TCA GCT CCC GGT CGC CGC	Protein	90
orf-37	57834-58304	Fwd:	ATG ACC GCG ACG AAT CCT	Regulatory	91
		Rev:	CTA GGC GGC GCG TCC CGC		92
orf-38	58440-60091	Fwd:	ATG AGC ACC ACG GCC GAG	Oxidoreductase	93
		Rev:	TCA GCC GCG CGC CGA CGG		94
orf-39	60092-60622	Fwd:	ATG ACC CTG GAG GCC TAC	Regulatory	95
		Rev:	TCA TGC GGG GCT CCC GGT		96
orf-40	60940-62020	Fwd:	GTG AAA AGT GAC TCT GCC	Regulatory	97
		Rev:	TCA ACG GCG AGT TGG CTG		98
orf-41	62045-62899	Fwd:	GTG ACC ACG AAC ACC ATC	Regulatory	99
		Rev:	TCA CCC GCG ATC TCG ATC		100
orf-42	62788-63164	Fwd:	(partial ORF)	p450	101
		Rev:	TCA CCT CGC CGT ACT CAC	hydroxylase	102

5

Surprisingly, sequence analysis failed to reveal any gene that resembles a polyketide synthase. The C-1027 open reading frames, however, encode polypeptides exhibiting a wide variety of enzymatic activities (*e.g.*, epoxide hydrase, monooxygenase, oxidoreductase, P-450 hydroxylase, *etc.*). The isolated C-1027 gene cluster can be used to

synthesize C-1027 enediyne antibiotics and/or analogues thereof. The C-1027 gene cluster can be modified and/or augmented to increase C-1027 and/or C-1027 analogue production.

Alternatively, various components of the C-1027 gene cluster can be used to synthesize and/or chemically modify a wide variety of metabolites. Thus, for example, ORF 6 (C-methyltransferase) can be used to methylate a carbon, while ORF 12, an epimerase, can be used to change the conformation of a sugar. The ORFs can be combined in their native configuration or in modified configurations to synthesize a wide variety of biomolecules/metabolites. Thus, for example, various combinations of C-1027 open reading frames can be used to synthesize an enediyne core, to synthesize a deoxy sugar, to synthesize a β-amino acid, to make a benzoxazolinate, *etc* (*see, e.g.,* Figures 2, 3, and 4).

The native C-1027 gene cluster ORFs can be re-ordered, modified, and combined with other biosynthetic units (*e.g.* polyketide synthases (PKSs) or catalytic domains thereof and/or non-ribosomal polypeptide synthetases (NRPSs) or catalytic domains thereof) to produce a wide variety of molecules. Large chemical libraries can be produced and then screened for a desired activity.

The C-1027 gene cluster also includes a number of drug resistance genes (*see, e.g.,* Table 2) that confer resistance to C-1027 and/or metabolites involved in C-1027 biosynthesis thereby permitting the cell to complete the enediyne biosynthesis. These resistance genes can be used to confer enediyne resistance on a cell lacking such resistance or to augment the enediyne resistance of a cell that does tolerate enediynes. Such cells can be used to produce high levels of enediynes and/or enediyne metabolites, and/or enediyne analogues.

Table III. C-1027 cluster drug resistance genes.

ORF	Protein	Mechanism
ORF 9:	CagA apoprotein	Drug sequestering
ORF 2:	SgcB transmembrane efflux protein	Drug exporting
ORF 27	Transmembrane transport protein	Drug exporting
ORF 0	Na^+/H^+ transporter	Drug exporting
ORF -1	ABC transport (C-terminus)	Drug exporting
ORF -2	Glycerol phosphate transporter	Drug exporting
ORF -1	UvrA-like protein (N-terminus)	DNA repairing

I. Isolation, preparation, and expression of C-1027 nucleic acids.

The C-1027 gene cluster nucleic acids can be isolated, optionally modified, and inserted into a host cell to create and/or modify a metabolic (biosynthetic) pathway and thereby enable that host cell to synthesize and/or modify various metabolites. Alternatively the C-1027 gene cluster nucleic acids can be expressed in the host cell and the encoded C-1027 polypeptide(s) recovered for use as chemical reagents, *e.g.* in the *ex vivo* synthesis and/or chemical modification of various metabolites. Either application typically entails insertion of one or more nucleic acids encoding one or more isolated and/or modified C-1027 enediyne open reading frames in a suitable host cell. The nucleic acid(s) are typically in an expression vector, a construct containing control elements suitable to direct expression of the C-1027 polypeptides. The expressed C-1027 polypeptides in the host cell then act as components of a metabolic/biosynthetic pathway (in which case the synthetic product of the pathway is typically recovered) or the C-1027 polypeptides themselves are recovered. Using the sequence information provided herein, cloning and expression of C-1027 nucleic acids can be accomplished using routine and well known methods.

A) C-1027 nucleic acids.

The nucleic acids comprising the C-1027 gene cluster are identified in Tables I and are listed in the sequence listing provided herein. In particular, Table 1 identifies genes and functions of open reading frames (ORFs) in the C-1027 enediyne biosynthesis gene cluster and identifies primers suitable for the amplification/isolation of any one or more of the C-1027 open reading frames. Of course, using the sequence information provided herein, other primers suitable for amplification/isolation of one or more C-1027 open reading frames can be determined according to standard methods well known to those of skill in the art (*e.g.* using Vector NTI Suite™, InforMax, Gaithersberg, MD, USA).

Typically such amplifications will utilize the DNA or RNA of an organism containing the requisite genes (*e.g. Streptomyces globisporus*) as a template. Typical amplification conditions include the following PCR temperature program: initial denaturing at 94°C for 5 min, 24-36 cycles of 45 sec at 94°C, 1 min at 60°C, 2 min at 72°C, followed by additional 7 min at 72°C. One of skill will appreciate that optimization of such a protocol, *e.g.* to improve yield, *etc.* is routine (*see, e.g.,* U.S. Patent No. 4,683,202; Innis (1990) *PCR*

Protocols A Guide to Methods and Applications Academic Press Inc. San Diego, CA, *etc*).

In addition, primer may be designed to introduce restriction sites and so facilitate cloning of the amplified sequence into a vector.

In one embodiment, this invention provides nucleic acids for the recombinant expression of an enediyne (*e.g.* a C-1027 enediyne or an analogue thereof). Such nucleic acids include isolated gene cluster(s) comprising open reading frames encoding polypeptides sufficient to direct the assembly of the enediyne. In other embodiments of this invention, the C-1027 open reading frames may be unchanged, but the control elements (*e.g.* promoters, enhancers, *etc.*) may be modified. In still other embodiments, the nucleic acids may encode selected components (*e.g.* one or more C-1027 or modified C-1027 open reading frames) and/or may optionally contain other heterologous biosynthetic elements including, but not limited to polyketide synthase (PKS) and/or non-ribosomal polypeptide synthetase (NRPS) modules or enzymatic domains.

Such variations may be introduced by design, for example to modify a known molecule in a specific way, *e.g.* by replacing a single substituent of the enediyne with another, thereby creating a derivative enediyne molecule of predicted structure. Alternatively, variations can be made randomly, for example by making a library of molecular variants of a known enediyne by systematically or haphazardly replacing one or open reading frames in the biosynthetic pathway. Production of alternative/modified enediyne, and hybrid enediyne PKSs and/or NRPSs and hybrid systems is described below.

Using the information provided herein other approaches to cloning the desired sequences will be apparent to those of skill in the art. For example, the enediyne, and/or optionally PKS and/or NRPS modules or enzymatic domains of interest can be obtained from an organism that expresses such, using recombinant methods, such as by screening cDNA or genomic libraries, derived from cells expressing the gene, or by deriving the gene from a vector known to include the same. The gene can then be isolated and combined with other desired biosynthetic elements using standard techniques. If the gene in question is already present in a suitable expression vector, it can be combined *in situ,* with, e.g., other PKS subunits, as desired. The gene of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence can be assembled from overlapping oligonucleotides prepared by standard methods and assembled

-21-

into a complete coding sequence (*see, e.g.,* Edge (1981) *Nature* 292:756; Nambair *et al.* (1984) *Science* 223: 1299; Jay *et al.* (1984) *J. Biol. Chem.* 259:6311). In addition, it is noted that custom gene synthesis is commercially available (*see, e.g.* Operon Technologies, Alameda, CA).

5   Examples of such techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel (1989) *Guide to Molecular Cloning Techniques, Methods in Enzymology 152* Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY; Ausubel (19

10   1994) *Current Protocols in Molecular Biology,* Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., U.S. Patent 5,017,478; and European Patent No. 0,246,864.

B)   Expression of f C-1027 open reading frames.

The choice of expression vector depends on the sequence(s) that are to be

15   expressed. Any transducible cloning vector can be used as a cloning vector for the nucleic acid constructs of this invention. However, where large clusters are to be expressed, it phagemids, cosmids, P1s, YACs, BACs, PACs, HACs or similar cloning vectors be used for cloning the nucleotide sequences into the host cell. Phagemids, cosmids, and BACs, for example, are advantageous vectors due to the ability to insert and stably propagate therein

20   larger fragments of DNA than in M13 phage and lambda phage, respectively. Phagemids which will find use in this method generally include hybrids between plasmids and filamentous phage cloning vehicles. Cosmids which will find use in this method generally include lambda phage-based vectors into which cos sites have been inserted. Recipient pool cloning vectors can be any suitable plasmid. The cloning vectors into which pools of

25   mutants are inserted may be identical or may be constructed to harbor and express different genetic markers (see, *e.g.,* Sambrook *et al., supra*). The utility of employing such vectors having different marker genes may be exploited to facilitate a determination of successful transduction.

In preferred embodiments of this invention, vectors are used to introduce C-

30   1027 biosynthesis genes or gene clusters into host (*e.g. Streptomyces*) cells. Numerous vectors for use in particular host cells are well known to those of skill in the art. For example described in Malpartida and Hopwook, (1984) *Nature,* 309:462-464; Kao *et al.,*

-22-

(1994), *Science*, 265: 509-512; and Hopwood *et al.*, (1987) *Methods Enzymol.*, 153:116-166 all describe vectors for use in various *Streptomyces* hosts.

In one preferred embodiment, *Streptomyces* vectors are used that include sequences that allow their introduction and maintenance in *E. coli*. Such *Streptomyces/E.*

5   *coli* shuttle vectors have been described (*see*, for example, Vara *et al.*, (1989) *J. Bacteriol.*, 171:5872-5881; Guilfoile & Hutchinson (1991) *Proc. Natl. Acad. Sci. USA*, 88: 8553-8557.)

The wildtype and/or modified C-1027 enediyne open reading frame(s) of this invention, can be inserted into one or more expression vectors, using methods known to those of skill in the art. Expression vectors will include control sequences operably linked to

10   the desired open reading frame. Suitable expression systems for use with the present invention include systems that function in eucaryotic and/or prokaryotic host cells. However, as explained above, prokaryotic systems are preferred, and in particular, systems compatible with *Streptomyces spp.* are of particular interest. Control elements for use in such systems include promoters, optionally containing operator sequences, and ribosome

15   binding sites. Particularly useful promoters include control sequences derived from enediyne, and/or PKS, and/or NRPS gene clusters. Other promoters (*e.g. ermE** as illustrated in Example 1) are also suitable. Other bacterial promoters, such as those derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) and maltose, will also find use in the present constructs. Additional examples include promoter sequences derived from

20   biosynthetic enzymes such as tryptophan (*trp*), the beta -lactamase (*bla*) promoter system, bacteriophage lambda PL, and T5. In addition, synthetic promoters, such as the *tac* promoter (U.S. Patent 4,551,433), which do not occur in nature also function in bacterial host cells. In *Streptomyces*, numerous promoters have been described including constitutive promoters, such as *ErmE* and *TcmG* (Shen and Hutchinson, (1994) *J. Biol. Chem.* 269: 30726-30733),

25   as well as controllable promoters such as *actI* and *actIII* (Pleper *et al.*, (1995) *Nature*, vol. 378: 263-266; Pieper *et al.*, (1995) *J. Am. Chem. Soc.*, 117: 11373-11374; and Wiesmann *et al.*, (1995) *Chem. & Biol.* 2: 583-589).

Other regulatory sequences may also be desirable which allow for regulation of expression of the enediyne open reading frame(s) relative to the growth of the host cell.

30   Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

-23-

Selectable markers can also be included in the recombinant expression vectors. A variety of markers are known which are useful in selecting for transformed cell lines and generally comprise a gene whose expression confers a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium. Such markers include, for example, genes that confer antibiotic resistance or sensitivity to the plasmid.

The various enediyne cluster open reading frames, and/or PKS, and/or NRPS clusters or subunits of interest can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, *e.g.*, a single promoter. The various open reading frames can include flanking restriction sites to allow for the easy deletion and insertion of other open reading frames so that hybrid synthetic pathways can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such as site-directed mutagenesis and PCR.

Methods of cloning and expressing large nucleic acids such as gene clusters, including PKS- or NRPS-encoding gene clusters, in cells including *Streptomyces* are well known to those of skill in the art (*see, e.g.*, Stutzman-Engwall and Hutchinson (1989) *Proc. Natl. Acad. Sci. USA*, 86: 3135-3139; Motamedi and Hutchinson (1987) *Proc. Natl. Acad. Sci. USA*, 84: 4445-4449; Grim *et al.* (1994) *Gene*, 151: 1-10; Kao *et al.* (1994) *Science*, 265: 509-512; and Hopwood *et al.* (1987) *Meth. Enzymol.*, 153: 116-166). In some examples, nucleic acid sequences of well over 100kb have been introduced into cells, including prokaryotic cells, using vector-based methods (*see*, for example, Osoegawa *et al.*, (1998) *Genomics*, 52: 1-8; Woon *et al.*, (1998) *Genomics*, 50: 306-316; Huang *et al.*, (1996) *Nucl. Acids Res.*, 24: 4202-4209). In addition, the cloning and expression of C-1027 enediyne is illustrated in Example 1.

C)   Host cells.

The vectors described above can be used to express various protein components of the enediyne, and/or enediyne shunt metabolites, and/or other modified metabolites for subsequent isolation and/or to provide a biological synthesis of one or more desired biomolecules (*e.g.* C-1027 and/or a C-1027 analogue, *etc.*). Where one or more proteins of the enediyne biosynthetic gene cluster are expressed (*e.g.* overexpressed) for subsequent isolation and/or characterization, the proteins are expressed in any prokaryotic or

-24-

eukaryotic cell suitable for protein expression. In one preferred embodiment, the proteins are expressed in *E. coli.*

Host cells for the recombinant production of the subject enediynes, enediyne metabolites, shunt metabolites, *etc.* can be derived from any organism with the capability of

5   harboring a recombinant enediyne gene cluster and/or subset thereof. Thus, the host cells of the present invention can be derived from either prokaryotic or eucaryotic organisms. Preferred host cells are those of species or strains (*e.g.* bacterial strains) that naturally express enediynes. Such host cells include, but are not limited to *Actinomycetes, Actinoplanetes,* and *Streptomycetes, Actinomadura, Micromonospra,* and the like.

10   Particularly preferred host cells include, but are not limited to *Streptomyces globisporus, Streptomyces lividans, Streptomyces coelicolor, Micromonospora echinospora* spp. *calichenisis, Actinomadura verrucosopora, Micromonospora chersina, Streptomyces carzinostaticus, and Actinomycete* L585-6. Other suitable host cells include, but are not limited to *S. verticillis S. ambofaciens, S. avermitilis, S. azureus, S. cinnamonensis, S.*

15   *coelicolor, S. curacoi, S. erythraeus, S. fradiae, S. galilaeus, S. glaucescens, S. hygroscopicus, S. lividans, S. parvulus, S. peucetius, S. rimosus, S. roseofulvus, S. thermotolerans,* and *S. violaceoruber* (*see, e.g.,* Hopwood and Sherman (1990) *Ann. Rev. Genet.* 24: 37-66; *O'Hagan* (1991) *The Polyketide Metabolites,* Ellis Horwood Limited, *etc.*).

20   In certain embodiments, a eukaryotic host cell is preferred (*e.g.* where certain glycosylation patterns are desired). Suitable eukaryotic host cells are well known to those of skill in the art. Such eukaryotic cells include, but are not limited to yeast cells, insect cells, plant cells, fungal cells, and various mammalian cells (*e.g.* COS, CHO HeLa cells lines and various myeloma cell lines).

25   **D)   Recovery of the expression product.**

Recovery of the expression product (*e.g.,* enediyne, enediyne analogue, enediyne biosynthetic pathway polypeptide, *etc.*) is accomplished according to standard methods well known to those of skill in the art. Thus, for example where enediyne biosynthetic gene cluster proteins are to be expressed and isolated, the proteins can be

30   expressed with a convenient tag to facilitate isolation (*e.g.* a His_6) tag. Other standard protein purification techniques are suitable and well known to those of skill in the art (*see,*

e.g., (Quadri *et al.* (1998) *Biochemistry* 37: 1585-1595; Nakano *et al.* (1992) *Mol. Gen. Genet.* 232: 313-321, *etc.*).

   Similarly where components (*e.g.* enediyne biosynthetic cluster orfs) are used to synthesize and/or modify various biomolecules (*e.g.* enediynes, enediyne analogues, shunt metabolites, *etc.*) the desired product and/or shunt metabolite(s) are isolated according to standard methods well know to those of skill in the art (*see, e.g.,* Carreras and Khosla (1998) *Biochemistry* 37: 2084-2088, Deutscher (1990) *Methods in Enzymology Volume 182: Guide to Protein Purification*, M. Deutscher, ed. *etc.*).

II. Use of C-1027 open reading frames in directed biosynthesis.

   Elements (*e.g.* open reading frames) of the C-1027 biosynthetic gene cluster and/or variants thereof can be used in a wide variety of "directed" biosynthetic processes (*i.e.* where the process is designed to modify and/or synthesize one or more particular preselected metabolite(s)). Essentially the entire C-1027 gene cluster can be used to synthesize a C-1027 enediyne and/or a C-1027 enediyne analogue. Individual C-1027 cluster open reading frames can be used to perform chemically modifications on particular substrates and/or to synthesize various metabolites. Thus, for example, ORF 6 (C-methyltransferase can be used to methylate a carbon), while ORF 7 (N-methyltransferase) can be used to methylate a nitrogen. ORF 12, and epimerase, can be used to change the conformation of a sugar, and ORF 8 (an amino transferase) can be used to aminate a suitable substrate. Similarly, combinations of C-1027 open reading frames can be used to direct the synthesis of various metabolites (*e.g.* β-amino acids, deoxysugars, benzoxazolinates, and the like). These examples, are merely illustrative. One of skill in the art, utilizing the information provided here, can perform literally countless chemical modifications and/or syntheses using either "native" enediyne biosynthesis metabolites as the substrate molecule, or other molecules capable of acting as substrates for the particular enzymes in question. Other substrates can be identified by routine screening. Methods of screening enzymes for specific activity against particular substrates are well known to those of skill in the art.

   The biosyntheses can be performed *in vivo*, *e.g.* by providing a host cell comprising the desired C-1027 gene cluster open reading frames and/or *in vivo*, *e.g.,* by providing the polypeptides encoded by the C-1027 gene cluster ORFs and the appropriate substrates and/or cofactors.

A)   Synthesis of enediynes and enediyne analogues.

In one embodiment, this invention provides for the synthesis of C-1027 enediynes and/or C-1027 analogues or derivatives. In a preferred embodiment, this is accomplished by providing a cell comprising a C-1027 gene cluster and culturing the cell under conditions whereby the desired enediyne or enediyne analogue is synthesized. The cell can be a cell that does not normally synthesize an enediyne and the entire gene cluster can be transfected into the cell. Alternatively, a cell that typically synthesizes enediynes can be utilized and all or part of the C-1027 gene cluster can be introduced into the cell.

Enediyne derivatives/analogues can be produced by varying the order of, or kind of, gene cluster subunits present in the cell, and/or by changing the host cell (*e.g.* to a eukaryotic cell that glycosylates the biosynthetic product), and/or by providing altered metabolites (*e.g.* adding exogenous aglycones to a host that carries a gene cassette of the deoxysugar biosynthesis and glycosylation genes for the production of glycosylated metabolites), *etc.*

In certain embodiments, the host cell need not be transfected with an entire C-1027 gene cluster. Rather, various components of a C-1027 gene cluster can be altered within a cell already harboring a C-1027 cluster. By varying or adding various biosynthetic open reading frames, C-1027 enediyne variants can be produced.

The use of standard techniques of molecular biology (gene disruption, gene replacement, gene supplement) can be used to modulate and/or otherwise alter enediyne and/or other metabolite (*e.g.* shunt metabolite) production in an organism that naturally synthesizes an enediyne (*e.g.* S. globisporus) or an organism that is modified to synthesize an enediyne.

In addition, or alternatively, control sequences that alter the expression of various open reading frames can be introduced that alter the amount and/or timing of enediyne production. Thus, for example, by placing particular C-1027 open reading frames under control of a constitutive promoter (*ermE**) C-1027 production was increased by as much as 4-fold (*see, e.g.* Table 3 and Example 1).

Table 3. Alteration of C-1027 production by engineering the C-1027 biosynthesis gene cluster.

Strain	Yield (%)

WT	100
WT/pKC1139	100
WT/*ermE**/ORF 2	>150
WT/ORF 9	>100
WT/*ermE**/ORF 9	<10
WT/ORF 10, 11	>100
WT/*ermE**/ORF 10, 11	>100
WT/ ORF 9, 10, 11	>400

ORF 2: transmembrane eflux protein; ORF 9: CagA apoprotein; ORF 10: TDP-glucose synthase; ORF 11; Hydroxylase/halogenase

Where enediyne analogues are synthesized, it will often prove desirable to
5   assay them for biological activity. Such assays are well know to those of skill in the art.
One such assay is illustrated in Example 1. Briefly, this example depicts an assay of
antibacterial activity against *M. luteus* as described by Hu *et al.* (1988) *J. Antibiot.* 41: 1575-
1579). Other suitable assays for enediyne activity will be known to those of skill in the art.

B) Use of C-1027 open reading frames to synthesize an enediyne core.

10   The C-1027 open reading frames described herein, or variants thereof, can be
used to synthesize an enediyene core, *e.g.*, from a fatty acid precursor. One such synthetic
pathway is illustrated in Figure 4. This reaction scheme utilizes ORF 17 (epoxide hydrase),
ORF 20 (monooxygenase), ORF 21 (iron-sulfur flavoprotein), ORF 29 (P-450 hydroxylase,
ORF 30 (oxidoreductase), ORF 32 (oxidoreductase), ORF 35 (proline oxidase), and ORF 38
15   (P-450 hydroxylase) to synthesize anenediyne core.

This synthetic pathway, is not considered limiting, but merely illustrative.
Using this as a model, one of ordinary skill in the art can design numerous other synthetic
schemes to produce enediyne cores and/or core variants.

C) Use of C-1027 open reading frames to synthesize deoxy sugars.

20   The biosynthesis of various deoxy sugars (*e.g.*, deoxyhexoses) typically share
a common key intermediate --4-keto-6-deoxyglucose nucleoside diphosphate or its analogs,
whose formation from glucose nucleoside diphosphate is catalyzed by the NGDH enzyme,
an NAD^+-dependent oxidoreductase (Liu and Thorson (1994) *Ann. Rev. Microbiol.* 48: 223-
256; Piepersberg (1997) pp. 81-163. In *Biotechnology of antibiotics,* 2nd ed. W. R. Strohl

(ed). Marcel Dekker, New York.). Similarly, the C-1027 gene cluster includes an NAGDH enzyme which can be exploited to synthesize a variety of deoxy sugars.

One illustrative synthetic pathway is shown in Figure 2. This biosynthetic scheme utilizes ORF 10 (dNDP-glucose synthase), ORF 1 (dNDP-glucose dehydratase),
5 ORF 12 (epimerase), ORF 8 (aminotransferase), ORF 6 (C-methyltransferase), ORF 7 (N-methyltransferase) and ORF 19 (glycosyl transferase).

This synthetic pathway, is not considered limiting, but merely illustrative. Using this as a model, one of ordinary skill in the art can design numerous other synthetic schemes to produce various deoxy sugars.

10 **D)   Use of C-1027 open reading frames to synthesize β-amino acids.**

In still another embodiment, C-1027 biosynthetic polypeptides can be used in the biosynthesis of β-amino acids. One illustrative synthetic pathway is shown in Figure 3A. This biosynthetic scheme utilizes ORF 4 (hydroxylase), ORF 11 (hydroxylase/halogenase), ORF 24 (aminomutase), ORF 23 (type II NRPS condensation enzyme), ORF 25 (type II
15 NRPS adenylation enzyme), and ORF 26 (type II peptidyl carrier protein).

Again, this synthetic pathway, is not considered limiting, but merely illustrative. Using this as a model, one of ordinary skill in the art can design numerous other synthetic schemes to produce other beta amino acids.

E)   Use of C-1027 open reading frames to synthesize benzoxazolinates.

20 The C-1027 open reading frames can also be used to synthesize a benzoxazolinate. One illustrative synthetic pathway is shown in Figure 3B. This biosynthetic scheme utilizes ORF 15 (anthranilate synthase I, ORF 16 (anthranilate synthase II), ORF 4 (phenol hydroxylase/chlorophenol-4-monooxygenase), ORF 11 (Hydroxylase/Halogenase), ORF 28 (O-methyltransferase), ORF 3 (coenzyme F390
25 synthetase, ORF 14 (coenzyme F390 synthetase), and ORF 13 (O-acyltransferase). Again, this synthetic pathway, is not considered limiting, but merely illustrative. Using this as a model, one of ordinary skill in the art can design numerous other synthetic schemes to produce other beta amino acids.

III. Generation of chemical diversity.

In addition to the directed modification and/or biosynthesis of various metabolites as described above, the C-1027 biosynthetic gene cluster open reading frames can be utilized, by themselves or in combination with other biosynthetic subunits (*e.g.* NRPS and/or PKS modules and/or enzymatic domains of other PKS and/or NRPS systems) to produce a wide variety of compounds including, but not limited to various enediyne or enediyne derivatives, various polyketides, polypeptides, polyketide/polypeptide hybrids, various thiazoles, various sugars, various methylated polypeptides/polyketides, and the like.

As with the directed production of various metabolites described above, such compounds can be produced, *in vivo* or *in vitro,* by catalytic biosynthesis, *e.g.,* using large, enediyne cluster units and/or modular PKSs, NRPSs, and hybrid PKS/NRPS systems. In a preferred embodiment large combinatorial libraries of cells harboring various megasynthetases can be produced by the random or directed modification of particular pathways and then selected for the production of a molecule or molecules of interest. It will be appreciated that, in certain embodiments, such libraries of megasynthetases/modified pathways, can be used to generate large, complex combinatorial libraries of compounds which themselves can be screened for a desired activity.

Such combinatorial libraries can be created by the deliberate modification/variation of selected biosynthetic pathways and/or by random/haphazard modification of such pathways.

A) Directed engineering of novel synthetic pathways.

In numerous embodiments of this invention, novel polyketides, polypeptides, and combinations thereof are created by modifying the entediyne gene cluster ORFs and/or known PKSs, and/or NRPSs so as to introduce variations into metabolites synthesized by the enzymes. Such variations may be introduced by design, for example to modify a known molecule in a specific way, *e.g.* by replacing a single monomeric unit within a polymer with another, thereby creating a derivative molecule of predicted structure. Such variations can also be made by adding one or more modules or enzymatic domains to a known PKS or NRPS or enediyne cluster, or by removing one or more module from a known PKS or NRPS.

Using any of these methods, it is possible to introduce PKS domains, NRPS domains, and entediyne domains into a megasynthetase. Mutations can be made to the

-30-

native enediyne, and/or NRPS, and/or PKS subunit sequences and such mutants used in place of the native sequence, so long as the mutants are able to function with other subunits (domains) in the synthetic pathway. Such mutations can be made to the native sequences using conventional techniques such as by preparing synthetic oligonucleotides including the mutations and inserting the mutated sequence into the gene encoding a NRPS and/or PKS subunit using restriction endonuclease digestion. (*see, e.g.,* Kunkel, (1985) *Proc. Natl. Acad. Sci. USA* 82: 448; Geisselsoder *et al.* (1987) *BioTechniques* 5: 786). Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) which hybridizes to the native nucleotide sequence (generally cDNA corresponding to the RNA sequence), at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located (Zoller and Smith (1983) *Meth, Enzymol.* 100: 468). Primer extension is effected using DNA polymerase, the product cloned and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Selection can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations (*see, e.g.,* Dalbie-McFarland *et al.* (1982) *Proc. Natl. Acad. Sci USA* 79:6409). PCR mutagenesis will also find use for effecting the desired mutations.

B) Random modification of enediyne pathways.

In another embodiment, variations can be made randomly, for example by making a library of molecular variants (*e.g.* of a known enediyne) by randomly mutating one or more elements of the subject gene cluster or by randomly replacing one or more open reading frames in a gene cluster with one or more of alternative open reading frames.

The various open reading frames can be combined into a single multi-modular enzyme, thereby dramatically increasing the number of possible combinations obtained using these methods. These combinations can be made using standard recombinant or nucleic acid amplification methods, for example by shuffling nucleic acid sequences encoding various modules or enzymatic domains to create novel arrangements of the sequences, analogous to DNA shuffling methods described in Crameri *et al.* (1998) *Nature* 391: 288-291, and in U.S. Patents 5,605,793 and in 5,837,458. In addition, novel combinations can be made in vitro, for example by combinatorial synthetic methods. Novel molecules or molecule libraries, can be screened for any specific activity using standard methods.

-31-

Random mutagenesis of the nucleotide sequences obtained as described above can be accomplished by several different techniques known in the art, such as by altering sequences within restriction endonuclease sites, inserting an oligonucleotide linker randomly into a plasmid, by irradiation with X-rays or ultraviolet light, by incorporating incorrect nucleotides during in vitro DNA synthesis, by error-prone PCR mutagenesis, by preparing synthetic mutants or by damaging plasmid DNA in vitro with chemicals. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, agents which damage or remove bases thereby preventing normal base-pairing such as hydrazine or formic acid, analogues of nucleotide precursors such as nitrosoguanidine, 5-bromouracil, 2-aminopurine, or acridine intercalating agents such as proflavine, acriflavine, quinacrine, and the like. Generally, plasmid DNA or DNA fragments are treated with chemicals, transformed into *E. coli* and propagated as a pool or library of mutant plasmids.

Large populations of random enzyme variants can be constructed in vivo using "recombination-enhanced mutagenesis." This method employs two or more pools of, for example, 10^6 mutants each of the wild-type encoding nucleotide sequence that are generated using any convenient mutagenesis technique, described more fully above, and then inserted into cloning vectors.

C) Incorporation and/or modification of non-C-1027 cluster elements.

In either the directed or random approaches, nucleic acids encoding novel combinations of gene cluster ORFs are introduced into a cell. In one embodiment, nucleic acids encoding one or more enediyne synthetic cluster ORFS and/or PKS and/or NRPS domains are introduced into a cell so as to replace one or more domains of an endogenous gene cluster within a cell. Endogenous gene replacement can be accomplished using standard methods, such as homologous recombination. Nucleic acids encoding an entire enediyne, enediyne ORF, PKS, NRPS, or combination thereof can also be introduced into a cell so as to enable the cell to produce the novel enzyme, and, consequently, synthesize the novel polymer. In a preferred embodiment, such nucleic acids are introduced into the cell optionally along with a number of additional genes, together called a 'gene cluster,' that influence the expression of the genes, survival of the expressing cells, etc. In a particularly preferred embodiment, such cells do not have any other enediyne and/or PKS- and/or NRPS-encoding genes or gene clusters, thereby allowing the straightforward isolation of the molecule(s) synthesized by the genes introduced into the cell.

Furthermore, the recombinant vector(s) can include genes from a single enediyne and/or PKS and/or NRPS gene cluster, or may comprise hybrid replacement PKS gene clusters with, *e.g.,* a gene for one cluster replaced by the corresponding gene from another gene cluster. For example, it has been found that ACPs are readily interchangeable among different synthases without an effect on product structure. Furthermore, a given KR can recognize and reduce polyketide chains of different chain lengths. Accordingly, these genes are freely interchangeable in the constructs described herein. Thus, the replacement clusters of the present invention can be derived from any combination of PKS and/or NRPS gene sets that ultimately function to produce an identifiable polyketide.

Examples of hybrid replacement clusters include, but are not limited to, clusters with genes derived from two or more of the act gene cluster, the *whiE* gene cluster, frenolicin (*fren*), granaticin (*gra*), tetracenomycin (*tcm*), 6-methylsalicylic acid (6-msas), oxytetracycline (*otc*), tetracycline (*tet*), erythromycin (*ery*), griseusin (*gris*), nanaomycin, medermycin, daunorubicin, tylosin, carbomycin, spiramycin, avermectin, monensin, nonactin, curamycin, rifamycin and candicidin synthase gene clusters, among others. (For a discussion of various PKSs, *see, e.g.,* Hopwood and Sherman (1990) *Ann. Rev. Genet.* 24: 37-66; O'Hagan (1991) The Polyketide Metabolites, Ellis Horwood Limited.

A number of hybrid gene clusters have been constructed, having components derived from the *act, fren, tcm, gris* and *gra* gene clusters (see, e.g., U.S. Patent 5,712,146). Other hybrid gene clusters, as described above, can easily be produced and screened using the disclosure herein, for the production of identifiable polyketides, polypeptides or polyketide/polypeptide hybrids.

Host cells (*e.g. Streptomyces*) can be transformed with one or more vectors, collectively encoding a functional PKS/NRPS set, or a cocktail comprising a random assortment of enediyne ORFs and/or PKS and/or NRPS genes, modules, active sites, or portions thereof. The vector(s) can include native or hybrid combinations of enediyne ORFs, and/or PKS and/or NRPS subunits or cocktail components, or mutants thereof. As explained above, the gene cluster need not correspond to the complete native gene cluster but need only encode the necessary enediyne ORFs and/or PKS and/or NRPS components to catalyze the production of the desired product(s).

IV.   Variation of starter and/or extender units, and/or host cells.

In addition to varying the nucleic acids comprising the subject gene cluster, variations in the products produced by the gene cluster(s) can be obtained by varying the the host cell, the starter units and/or the extender units. Thus, for example different fatty acids can be utilized in the enediyne synthetic pathway resulting in different enediyne variants. Similarly different intermediate metabolites can be provided (*e.g.* endogenously produced by the host cell, or produced by an introduced herterologous construct, and/or supplied from an exogenous source (*e.g.* the culture media)). Similarly, varying the host cell can vary the resulting product(s). For example, a gene cassette carrying the enediyne biosynthesis genes can be introduced into a deoxysugar-synthesizing host for the production of glycosylated enediyne metabolites.

V.   Use of C-1027 resistance genes.

The antibiotic C-1027 and metabolites present in C-1027 biosynthesis are highly potent cytotoxins. Accordingly the biosynthesis of C-1027 is facilitated by the presence of one or more antibiotic (*e.g.* enediyne) resistance genes. Without being bound to a particular theory, it is believed that CagA and SgcB function cooperatively to provide resistance. It is believed that the C-1027 chromophore is first sequestered by binding to the preaproprotein CagA (ORF 9) to form a complex, which is then transported out of the cell by the efflux pump SgcB (ORF 2) and processed by removing the leader peptide to yield the chromoprotein. Other genes that appear to mediate resistance in the C-1027 biosynthesis gene cluster include a transmembrane transport protein (ORF 27), a Na^+/H^+ transporter (ORF 0), an ABC transporter (ORF -1, C-terminus), a glycerol phosphate transporter (ORF -2), and a UvrA-like protein (ORF -1, N-terminus) (*see, e.g.,* Table 2).

These ORFs and/or the polypeptides encoded by these ORFs can be utilized alone, or in combination with one or more other C-1027 ORFs to confer resistance to enediyne or enediyne metabolites on a cell. This is useful in a wide variety of contexts. For example, to increase production of enediynes. For example, it is believed that C-1027 resistance could be a limiting factor at the onset of C-1027 production. Provision of an extra copy of the plasmid-born *sgcB*, and overexpression of *sgcB* under the control of the constitutive *ermE** promoter resulted in increase of C-1027 production (see example 1).

In a therapeutic context, it is sometimes desirable to confer resistance on certain vulnerable cells. Thus, for example, where an enediyne is used as a

chemotherapeutic, transfection of vulnerable, but healthy cells (*e.g.* liver cells remote from the tumor site, stem cells, *etc.*) with vector(s) expressing the resistance gene(s) permits administration of the enediyne at a higher dosage with fewer adverse effects to the organism. Such approaches have been taken using the multi-drug resistance gene (MDR1) expressing p-glycoprotein.

In another embodiment vectors are provided containing one or more resistance genes of this invention under control of a constitutive and/or inducible promoter thereby providing a "ready-made" expression system suitable for the expression of an enediyne or enediyne metabolite at high concentration.

It is also noted that the resistance genes are expected to confer resistance to compounds other than enediynes. The resistance genes are expected to confer resistance to essentially any cytotoxic compound that can act as a substrate for the resistance gene(s) of this invention.

VI.   Kits.

In still another embodiment, this invention provides kits for practice of the methods described herein. In one preferred embodiment, the kits comprise one or more containers containing nucleic acids encoding one or more of the C-1027 biosynthesis gene cluster open reading frames. Certain kits may comprise vectors encoding the *sgc* gene cluster orfs and/or cells containing such vectors. The kits may optionally include any reagents and/or apparatus to facilitate practice of the methods described herein. Such reagents include, but are not limited to buffers, labels, labeled antibodies, bioreactors, cells, *etc.*

In addition, the kits may include instructional materials containing directions (*i.e.*, protocols) for the practice of the methods of this invention. Preferred instructional materials provide protocols utilizing the kit contents for creating or modifying C-1027 gene cluster and/or for synthesizing or modifying a molecule using one or more *sgc* gene cluster ORFs. While the instructional materials typically comprise written or printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this invention. Such media include, but are not limited to electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (e.g., CD ROM), and the like. Such media may include addresses to internet sites that provide such instructional materials.

-35-

The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1

5   **Genes for production of the enediyne antitumor antibiotic C-1027 in *Streptomyces globisporus* are clustered with the *cagA* gene that encodes the C-1027 apoprotein**

We have been studying the biosynthesis of C-1027 in *Streptomyces globisporus* C-1027 as a model for the enediyne family of antitumor antibiotics (Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188). C-1027 consists of a non-peptidic chromophore and
10   an apoprotein, CagA [also called C-1027AG (Otani *et al.* (1991) *Agri. Biol. Chem.* 55: 407-417)]. The C-1027 chromophore is extremely unstable in the protein-free state, the structure of which was initially deduced from an inactive but more stable degradation product (Minami *et al.* (1993) *Tetrahedron Lett.* 34: 2633-2636) and subsequently confirmed by spectroscopic analysis of the natural product (Yoshida *et al.* (1993) *Tetrahedron Lett.* 34:
15   2637-2640) (Fig. 1). While the absolute stereochemistry of the deoxysugar moiety was established by total synthesis (Iida *et al.* (1993) *Tetrahedron Lett.* 34: 4079-4082), the *8S, 9S, 13S* and *17R* configuration of the C-1027 chromophore were based only on computer modeling (Okuno *et al.* (1994) *J. Med. Chem.* 37: 2266-2273). Although no biosynthetic study has been carried out specifically on C-1027, the polyketide origin of the enediyne
20   cores has been implicated by feeding experiments with ^{13}C-labeled acetate for the neocarzinostatin chromophore A (Hensens *et al.* (1989) *J. Am. Chem. Soc.* 111: 3295-3299), dynemicin (Tokiwa *et al.* (1992) *J. Am. Chem. Soc.* 114: 4107-4110), and esperamicin (Lam *et al.* (1993) *J. Am. Chem. Soc.* 115: 12340-12345); and deoxysugar biosynthesis has been well characterized in actinomycetes (Liu and Thorson (1994) *Ann. Rev. Microbiol.* 48: 223-
25   256; Piepersberg (1997) pp. 81-163. In *Biotechnology of antibiotics,* 2nd ed. W. R. Strohl (ed). Marcel Dekker, New York). Given the structural similarity of C-1027 to the other enediyne cores and to deoxysugars found in other secondary metabolites, we decided to clone either a PKS or a deoxysugar biosynthesis gene as the first step of identifying the C-1027 gene cluster from *S. globisporus*.

30   Furthermore, the CagA apoprotein of C-1027 has been isolated, its amino acid sequence has been determined, and the corresponding *cagA* gene has been cloned and sequenced (Otani *et al.* (1991) *Agri. Biol. Chem.* 55: 407-417; Sakata *et al.* (1992) *Biosci.*

Biotech. Biochem. 56: 1592-1595). Since genes encoding secondary metabolite production in actinomycetes have invariably been found to be clustered in one region of the microbial chromosome (Hopwood (1997) *Chem. Rev.* 97: 2465-2497), we further reasoned that mapping the *cagA* gene with either a putative PKS gene, a deoxysugar biosynthesis gene, or

5   both to the same region of the *S. globisporus* chromosome should be viewed as strong evidence supporting the proposition that the cloned genes constitute the C-1027 biosynthesis gene cluster.

We report here the cloning and sequencing of two genes, *sgcA* (*Streptomyces globisporus* C-1027) and *sgcB*, that encode a dNDP-glucose 4,6-dehydratase (NGDH) and a

10   transmembrane efflux protein, respectively. The *sgcA,B* locus is indeed clustered with the *cagA* gene, leading to the localization of a 75-kb gene cluster from *S. globisporus*. The involvement of the cloned gene cluster in C-1027 biosynthesis was demonstrated by disrupting the *sgcA* gene to generate C-1027-nonproducing mutants and by complementing the *sgcA* mutants in vivo to restore C-1027 production. Our results, together with similar

15   effort in the Thorson laboratory on the calicheamicin gene cluster (Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188), represent the first cloning of a gene cluster for enediyne antitumor antibiotic biosynthesis.

Materials and methods.

Bacterial strains and plasmids.

20   *Escherichia coli* DH5α was used as a general host for routine subcloning (Sambrook *et al.* (1989) *Molecular cloning, a laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). *E. coli* XL 1-Blue MR (Stratagene, La Jolla, CA) was used as the transduction host for cosmid library construction. *E. coli* S17-1 was used as the donor host for *E. coli-S. globisporus* conjugation (Mazodier *et al.* (1989) *J. Bacteriol.*

25   171: 3583-3585). *Micrococcus luteus* ATCC9431 was used as the testing organism to assay the antibacterial activity of C-1027 (Hu *et al.* (1988) *J. Antibiot.* 41: 1575-1579). The pGEM-3zf, -5zf, and -7zf and pGEM-T vectors were from Promega (Madison, WI). *S. globisporus* strains and other plasmids in this study are listed in Table 3

30   **Table 3.** Strains and plasmids.

Strain or	Relevant Characteristics

plasmid	

S. globisporus

C-1027	Wild-type (Hu *et al.* (1988) *J. Antibiot.* 41: 1575-1579)
AF40	Mutant resulted from acriflavine treatment of *S. globisporus* C-1027, C-1027-nonproducing (Mao *et al.* (1997) *Chinese J. Biotechnol.* 13: 195-199)
AF44	Mutant resulted from acriflavine treatment of *S. globisporus* C-1027, C-1027-nonproducing (Mao *et al., supra*)
AF67	Mutant resulted from acriflavine treatment of *S. globisporus* C-1027, C-1027-nonproducing (Mao *et al., supra*)
SB1001	*sgcA*-disrupted mutant resulted from integration of pBS1012 into *S. globisporus* C-1027 Apr^R, C-1027-nonproducing
SB1002	*sgcA*-disrupted mutant resulted from integration of pBS1013 into *S. globisporus* C-1027 Apr^R, C-1027-nonproducing

Plasmids:

pOJ446	*E. coli-Streptomyces* shuttle cosmid, Apr^R (Bierman *et al.* (1992) *Gene,* 116: 43-
pOJ260	*E. coli* vector, non-replicating in *Streptomyces*, Apr^R (Bierman *et al. supra)*
pKC1139	*E. coli-Streptomyces* shuttle vector, rep^{TS}, Apr^R (Bierman *et al. supra)*
pWHM3	*E. coli-Streptomyces* shuttle vector, Th^R (Vara *et al.* (1989) *J. Bacteriol.* 171: 5872-5881)
pWHM79	*ermE** promoter in pGEM-3zf (Shen and Hutchinson (1996) *Proc. Natl. Acad. Sci. USA* 93: 6600-6604)
pBS1001	0.75-kb PCR product amplified from *S. globisporus* with type I PKS primers in pGEM-T
pBS1002	0.55-kb PCR product amplified from *S. globisporus* with NGDH gene primers in pGEM-T
pBS1003	0.73-kb PCR product amplified from pBS1005 with *cagA* primers in pGEM-T
pBS1004	pOJ446 *S. globisporus* genomic library cosmid
pBS1005	pOJ446 *S. globisporus* genomic library cosmid
pBS1006	pOJ446 *S. globisporus* genomic library cosmid
pBS1007	3.0-kb *Bam*HI fragment from pBS1005 in pGEM-3zf, *sgcA, sgcB*
pBS1008	4.0-kb *Bam*HI fragment from pBS1005 in pGEM-3zf, *cagA*
pBS1009	1.0-kb *Kpn*I truncated fragment of *sgcA* from pBS1007 in pGEM-3zf
pBS1010	0.75-kb *Sac*II/*Sph*I internal fragment of *sgcA* from pBS1009 in pGEM-5zf

pBS1011	0.75-kb *Sac*I/*Sph*I internal fragment of *sgcA* from pBS1010 in pGEM-3zf
pBS1012	0.75-kb *Eco*RI/*Hind*III internal fragment of *sgcA* from pBS1010 in pOJ260
pBS1013	0.75-kb *Eco*RI/*Hind*III internal fragment of *sgcA* from pBS1010 in pKC1139
pBS1014	2.0-kb *Eco*RI/*Sph*I fragment from pBS1007 in the *Sma*I/*Sph*I sites of pWHM79, *ermE**, *sgcA*
pBS1015	2.5-kb *Eco*RI/*Hind*III fragment from pBS1014 in pWHM3, *ermE**, *sgcA*
pBS1016	Self-ligation of the 5.2-kb *Kpn*I fragment from pBS1007
pBS1017	0.45-kb *Eco*RI/*Sac*I fragment from pWHM79 in *Eco*RI/*Sac*I sites of pBS1016, *ermE**, *sgcB*
pBS1018	2.5-kb EcoRI/HindIII fragment from pBS1017 in pKC1139, *ermE**, *sgcB*

Biochemicals and chemicals.

Ampicillin, apramycin, nalidixic acid, and thiostrepton were from Sigma (St. Louis, MO). Unless specified otherwise, restriction enzymes and other molecular biology

5   reagents were from standard commercial sources.

Media and culture conditions.

E. coli strains carrying plasmids were grown in Luria-Bertani (LB) medium and were selected with appropriate antibiotics. *S. globisporus* strains were grown on ISP-4 (Difco Laboratories, Detroit, MI) or R2YE at 28°C for sporulation and in TSB (Hopwood *et*

10   *al.* (1985) *Genetic manipulation of Streptomyces: a laboratory manual.* John Innes Foundation, Norwich, UK) supplemented with 5 mM MgCl$_2$ and 0.5% glycine at 28°C, 250 rpm for isolation of genomic DNA. For transformation, *S. globisporus* strains were grown in YEME (Hopwood *et al., supra.*) for preparation of protoplasts and on R2YE for protoplast regeneration. For conjugation, both the *E. coli* S17-1 donors and the *S. globisporus*

15   recipients (upon germination in TSB) were prepared in LB, and donors/recipients were grown on either ISP-4 medium with 0.05% yeast extract and 0.1% tryptone or AS-1 medium (Baltz (1980) *Dev. Ind. Microbiol.* 21: 43-54; Bierman *et al.* (1992) *Gene* 116: 43-69) at 30°C for isolation of exconjugants.

For C-1027 production, *S. globisporus* strains were grown either on R2YE or

20   ISP-4 agar medium at 28°C or in liquid medium by a two-stage fermentation. For liquid culture, the seed inoculum was prepared by inoculating 50 mL medium (consisting of 2% glycerol, 2% dextrin, 1% fish meal, 0.5% peptone, 0.2% (NH$_4$)$_2$SO$_4$, and 0.2% CaCO$_3$, pH 7.0) with an aliquot of spore suspension, incubating at 28°C, 250 rpm for two days. To a

fresh 50 mL of the same medium was then added the seed culture (5%), and incubation continued at 28°C, 250 rpm for three to six days (Hu et al. (1988) J. Antibiot. 41: 1575-1579). The fermentation supernatants were harvested by centrifugation (Eppendorf 5415C, 4°C, 10 min, 14,000 rpm) on day 3, 4 and 5, and assayed for their antibacterial activity

5   against M. luteus (Hu et al. (1988) J. Antibiot., 41: 1575-1579).

DNA isolation and manipulation.

Plasmid preparation and DNA extraction were carried out by using commercial kits (Qiagen, Santa Clarita, CA). Total S. globisporus DNA was isolated according to literature protocols (Hopwood et al. (1985) Genetic manipulation of

10   Streptomyces: a laboratory manual. John Innes Foundation, Norwich, UK; Rao et al. (1987) Methods Enzymol. 153: 166-198). Restriction endonuclease digestion and ligation followed standard methods (Sambrook et al. (1989) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). For Southern analysis, digoxigenin labeling of DNA probes, hybridization, and detection were performed according to the

15   protocols provided by the manufacturer (Boehringer Mannheim Biochemicals, Indianapolis, IN).

DNA sequencing.

Automated DNA sequencing was carried out on an ABI Prism 377 DNA Sequencer using the ABI Prism dye terminator cycle sequencing ready reaction kit and

20   AmpliTaq DNA polymerase FS (Perkin-Elmer/ABI, Foster City, CA). Sequencing service was provided by either the DBS Automated DNA Sequencing Facility, UC Davis, or Davis Sequencing Inc. (Davis, CA). Data were analyzed by ABI Prism Sequencing 2.1.1 software and the Genetics Computer Group program (Madison, WI).

Polymerase chain reaction (PCR).

25   Primers were synthesized at the Protein Structure Laboratory, UC Davis. PCR was carried out on a Gene Amp PCR System 2400 (Perkin-Elmer/ABI) with Taq polymerase and buffer from Promega. A typical PCR mixture consisted of 5 ng of S. globisporus genomic or plasmid DNA as template, 25 pmoles of each primers, 25 μM dNTP, 5% DMSO, 2 units of Taq polymerase, 1 x buffer, with or without 20% glycerol in a final

30   volume of 50 μL. The PCR temperature program was as follows: initial denaturing at 94°C

for 5 min, 24-36 cycles of 45 sec at 94°C, 1 min at 60°C, 2 min at 72°C, followed by additional 7 min at 72°C.

For type II PKS, the following two pairs of degenerate primers were used—5'-AGC TCC ATC AAG TCS ATG RTC GG-3' (forward, SEQ ID NO:103) / 5'-CC GGT GTT SAC SGC GTA GAA CCA GGC G-3' (reverse, SEQ ID NO:104) and 5'-GAC ACV GCN TGY TCB TCV-3' (forward, SEQ ID NO: 105)/5'-RTG SGC RTT VGT NCC RCT-3' (SEQ ID NO:106) (B, C+G+T; N, A+C+G+T; R, A+G; S, C+G; V, A+C+G; Y, C+T) (reverse) (Seow *et al.* (1997) *J. Bacteriol.,* 179: 7360-7368). No product was amplified under all conditions tested. For type I PKS, the following pair of degenerate primers were used—5'-GCS TCC CGS GAC CTG GGC TTC GAC TC-3' (forward, SEQ ID NO:107) / 5'-AG SGA SGA SGA GCA GGC GGT STC SAC-3' (S, G+C) (reverse, SEQ ID NO:108) (Kakavas *et al.* (1997) *J. Bacteriol.,* 179: 7515-7522). A distinctive product with the predicted size of 0.75 kb was amplified in the presence of 20% glycerol and cloned into pGEM-T according to the protocol provided by the manufacturer (Promega) to yield pBS1001.

For NGDH, the following pair of degenerate primers were used—5'-CS GGS GSS GCS GGS TTC ATC GG-3' (forward, SEQ ID NO:109) / 5'-GG GWR CTG GYR SGG SCC GTA GTT G-3' (R, A+G; S, C+G; W, A+T; Y, C+T) (reverse, SEQ ID NO:110) (Decker, *et al.* (1996) FEMS Lett., 141: 195-201). A distinctive product with the predicted size of 0.55 kb was amplified and cloned into pGEM-T to yield pBS1002.

For *cagA*, the following pair of primers, flanking its coding region, were used—5'-AG GTG GAG GCG CTC ACC GAG-3' (forward, SEQ ID NO:111)/5'-G GGC GTC AGG CCG TAA GAA G-3' (reverse, SEQ ID NO:112) (Sakata *et al.* (1992) *Biosci. Biotechnol. Biochem.,* 56: 159201595). A distinctive product with the predicted size of 0.73 kb was amplified from pBS1005 and cloned into pGEM-T to yield pBS1003.

Genomic library construction and screening.

S. globisporus genomic DNA was partially digested with *Mbo*I to yield a smear around 60 kb, as monitored by electrophoresis on a 0.3% agarose gel. This sample was dephosphorylated upon treatment with shrimp alkaline phosphatase and ligated into the *E. coli-Streptomyces* shuttle vector pOJ446 (Bierman *et al.* (1992) *Gene* 116: 43-69) that was prepared by digestion with *Hpa*I, shrimp alkaline phosphatase treatment, and additional digestion with *Bam*HI. The resulting ligation mixture was packaged with the Gigapack II

-41-

XL two-component packaging extract (Stratagene). The package mixture was transduced into *E. coli* XL 1-Blue MR. The transduced cells were spread onto LB plates containing apramycin (100 μg/mL) and incubated at 37°C overnight. The titer of the primary library was approximately 6,000 colony-forming units per μg of DNA. Restriction enzyme analysis
5   of twelve randomly selected cosmids confirmed that the average size of inserts was about 35 to 45 kb (Rao *et al.* (1987) *Meth. Enzymol.*, 153: 166-198).

To screen the genomic library, colonies from five LB plates containing apramycin (100 μg/mL, with approximately 2,000 colonies per plate) were transferred to nylon transfer membranes (Micro Separations, Inc., Westborough, MA) and screened by
10   colony hybridization with the PCR-amplified 0.55-kb NGDH fragment from pBS1002 as a probe. The positive cosmid clones were re-screened by PCR with primers for NGDH and confirmed by Southern hybridization (Sambrook *et al., supra.*). Further restriction enzyme mapping and chromosomal walking of these overlapping cosmids led to the genetic localization of the 75-kb *sgc* gene cluster, as represented by pBS1004, pBS1005, and
15   pBS1006 (Fig. 5A). A 3.0-kb *Bam*HI fragment from pBS1005 that hybridized to the NGDH probe was cloned into the same sites of pGEM-3zf to yield pBS1007. Similarly, a 4.0-kb *Bam*HI fragment from pBS1005 that hybridizes to the PCR-amplified 0.73-kb *cagA* probe from pBS1003 was cloned into the same sites of pGEM-3zf to yield pBS1008 (Fig. 5B).

Generation of *sgcA* mutants by insert-directed homologous recombination in *S.*
20   ***globisporus.***

A 1.0-kb *Kpn*I fragment from pBS1007, containing the C-terminal truncated *sgcA*, was subcloned into pGEM-3zf to yield pBS1009. An internal fragment of *sgcA* was moved sequentially as a 0.75-kb *Sac*II/*Sph*I fragment from pBS1009 into the same sites of pGEM-5zf to yield pBS1010 and as a 0.75-kb *Sac*I/*Sph*I fragment from pBS1010 into the
25   same sites of pGEM-3zf to yield pBS1011. The latter plasmid was digested with *Eco*RI and *Hind*III, and the resulting 0.75-kb *Eco*RI/*Hind*III fragment was cloned into the same sites of pOJ260 and pKC1139 (Bierman *et al.* (1992) Gene, 116: 43-69 to yield pBS1012 and pBS1013, respectively.

Introduction of pBS1012 and pBS1013 into *S. globisporus* was carried out by
30   either polyethyleneglycol (PEG)-mediated protoplast transformation (Hopwood *et al.* (1985) *Genetic manipulation of Streptomyces: a laboratory manual.* John Innes Foundation, Norwich, UK) or *E. coli-S. globisporus* conjugation (Bierman *et al.* (1992) *Gene* 116: 43-69;

Matsushima and Baltz (1996) *Microbiology* 142: 261-267; Matsushima *et al.* (1994) *Gene* 146: 39-45), methods for both of which were developed recently in our laboratory. In brief, for transformation, pBS1012 and pBS1013 were propagated in *E. coli* ET12567 (MacNeil *et al.* (1992) *Gene* 111: 61-68), and the resulting double strand plasmid DNA was denatured by

5 alkaline treatment (Ho and Chater (1997) *J. Bacteriol.* 179: 122-127). The latter DNA (5 μL) and 200 μL of 25% PEG 1000 in P buffer (Hopwood *et al. supra*) were sequentially added to 50 μL of *S. globisporus* protoplasts (10^9) in P buffer. The resulting suspension was mixed immediately and spread on R2YE plates. After incubation at 28°C for 16 to 20 hrs, the plates were overlaid with soft R2YE (0.7% agar) containing apramycin (100 μg/mL, final

10 concentration); incubation continued until colonies appeared (in 5 to 7 days). For conjugation, *E. coli* S17-1(pBS1012) or *E. coli* S17-1 (pBS1013) was grown to an OD_{600} of 0.3 to 0.4. Cells from a 20-mL culture were pelleted by centrifugation, washed in LB, and resuspended in 2 mL of LB as the *E. coli* donors. *S. globisporus* spores (10^3 to 10^9) were washed, resuspended in TSB, and incubated at 50°C for 10 min to activate germination.

15 After additional incubation at 37°C for 2 to 5 hrs, the spores were pelleted and resuspended in LB as the *S. globisporus* recipients. The donors (100 μL) and recipients (100 μL) were mixed and spread equally onto two modified ISP-4 or AS-1 plates supplemented freshly with 10 mM $MgCl_2$ (see Media and culture conditions). The plates were incubated at 28°C for 16 to 22 hrs. After removal of most of the *E. coli* S17-1 donors by washing the surface with

20 sterile water, the plates were overlaid with 3 mL of soft LB (0.7% agar) containing nalidixic acid (50 μg/mL, final concentration) and apramycin (100 μg/mL, final concentration) and incubated at 28°C until exconjugants appeared (in approximately 5 days).

Unlike pBS1012, which is a *Streptomyces* non-replicating plasmid, pBS1013 bears a temperature-sensitive *Streptomyces* replication origin (Bierman *et al.* (1992) *Gene*

25 116: 43-69; Muth *et al.* (1989) *Mol. Gen. Genet.* 219: 341-348) that is unable to replicate at temperatures above 34°C (Table 3), while the *S. globisporus* wild-type strain grows normally up to 37°C. Thus, spores of *S. globisporus* (pBS1013), from either the transformants or the exconjugants, were spread onto R2YE plates containing apramycin (100 μg/mL). The plates were incubated directly at 37°C, and mutants, resulting from single crossover homologous

30 recombination between pBS1013 and the *S. globisporus* chromosome, were readily obtained in 7 to 10 days. Alternatively, the plates were first incubated at 28°C for 2 days until pinpoint-size colonies became visible and then shifted to 37°C to continue incubation.

Mutants resulting from single crossover homologous recombination grew out of the original pinpoint-size colonies as easily distinguishable sectors in 7 to 10 days.

Construction of the *sgcA* and *sgcB* expression plasmids.

pBS1007 was digested with *Eco*RI, and made blunt-ended by treatment with

5 the Klenow fragment of DNA polymerase I. Upon additional digestion with *Sph*I, the resulting 2.0-kb blunt-ended *Sph*I fragment containing the intact *sgcA* gene was cloned into the *Sma*I/*Sph*I sites of pWHM79 (Shen *et al.* (1996) *Proc. Natl. Acad. Sci., USA,* 93: 6600-6604) to yield pBS1014. The latter was digested with *Eco*RI and *Hind*III, and the resulting 2.5-kb *Eco*RI/*Hind*III fragment was cloned into the same sites of pWHM3 (Vara *et al.*

10 (1989) *J. Bacteriol.* 171: 5872-5881) to yield pBS1015, in which the expression of *sgcA* is under the control of the *ermE** promoter (Bibb *et al.* (1994) *Mol. Microbiol.* 14: 533-545).

Alternatively, pBS1007 was digested with *Kpn*I, removing most of the *sgcA* gene, and the 5.2-kb *Kpn*I fragment was recovered and self-ligated to yield pBS1016. The *ermE** promoter was subcloned from pWHM79 (Shen *et al.* (1996) *Proc. Natl. Acad. Sci.,*

15 *USA,* 93: 6600-6604) as a 0.45-kb *Eco*RI/*Sac*I fragment and cloned into the same sites of pBS1016 to yield pBS1017. The latter was digested with *Eco*RI and *Hind*III, and the resulting 2.5-kb *Eco*RI/*Hind*III fragment was cloned into the same sites of pKC1139 to yield pBS1018, in which the expression of *sgcB* is under the control of the *ermE** promoter.

Determination of C-1027 production.

20 The production of C-1027 was detected by assaying its antibacterial activity against *M. luteus* (Hu *et al.* (1988) *J. Antibiot.* 41: 1575-1579). From liquid culture, fermentation supernant (180 µL) was added to stainless steel cylinders placed on LB plates pre-seeded with overnight *M. luteus* culture (0.01% vol/vol). From solid culture, a small square block (0.5 x 0.5 x 0.5 cm^3) of agar from either R2YE or ISP-4 medium was directly

25 placed on *M. luteus*-seeded LB plates. The plates were incubated at 37°C for 24 hrs, and C-1027 production was estimated by measuring the size of inhibition zones.

Nucleotide sequence accession number.

The nucleotide sequence reported here has been deposited in the GenBank database with the accession number AF201913.

Results.

No polyketide synthase gene was amplified by PCR from *S. globisporus*.

On the assumption that the C-1027 enediyne core is of polyketide origin, the PCR approach was adopted to screen *S. globisporus* for any putative PKS genes, although it

5   is far from certain *a priori* if the biosynthesis of the enediyne core invokes a PKS and, if so, whether the enediyne PKS will exhibit a type I or type II structural organization. PCR methods for cloning either type I or type II PKS genes have been developed, and these methods have proven to be very effective in cloning PKS genes from various polyketide-producing actinomycetes (Kakavas *et al.* (1997) *J. Bacteriol.* 179: 7515-7522; Seow *et al.*

10   (1997) *J. Bacteriol.* 179: 7360-7368). While no distinctive product was amplified under all conditions examined with both pairs of primers designed for type II PKS, a single product with the expected size of 0.75 kb was readily amplified by PCR from *S. globisporus* with primers designed for type I PKS, which was subsequently cloned (pBS1001). Intriguingly, sequence analysis of six randomly selected pBS1001 clones yielded an identical product—

15   indicative of a specific PCR amplification—the deduced amino acid sequence of which, however, showed no homology to known PKSs (data not shown), excluding the possibility of using PKS as a probe to identify the *sgc* biosynthesis gene cluster.

Cloning of a putative NGDH gene by PCR from *S. globisporus*.

The biosynthesis of various deoxyhexoses share a common key

20   intermediate—4-keto-6-deoxyglucose nucleoside diphosphate or its analogs—whose formation from glucose nucleoside diphosphate is catalyzed by the NGDH enzyme, an NAD^+-dependent oxidoreductase (Liu and Thorson (1994) *Ann. Rev. Microbiol.* 48: 223-256; Piepersberg (1997) pp. 81-163. In *Biotechnology of antibiotics,* 2nd ed. W. R. Strohl (ed). Marcel Dekker, New York). The PCR method was adopted to clone the putative

25   NGDH gene from *S. globisporus* with primers designed according to the homologous regions of various NGDH enzymes from actinomycetes (Decker *et al.* (1996) *FEMS Lett.* 141: 195-201), resulting in the amplification of a single product with the expected size of 0.55 kb (pBS1002). Sequence analysis of pBS1002 confirmed its identity as a part of a putative NGDH gene.

30   To clone the complete NGDH gene, an *S. globisporus* genomic library, constructed in the *E. coli-Streptomyces* shuttle vector pOJ446 (Bierman *et al.* (1992) *Gene*

116: 43-69; Rao *et al.* (1987) *Methods Enzymol.* 153: 166-198), was analyzed by Southern hybridization with the PCR-amplified 0.55-kb fragment from pBS1002 as a probe. Of the 10,000 colonies screened, 36 positive colonies were identified, 9 of which were confirmed by PCR to harbor the DGDH gene. Restriction enzyme mapping showed that all of them

5 contained a single 3.0-kb *Bam*HI fragment hybridizing to the NGDH probe. Additional chromosomal walking from this locus eventually led to the localization of the 75-kb *sgc* gene cluster, covered by 18 overlapping cosmids as represented by pBS1004, pBS1005, and pBS1006 (Fig. 5A). The 3.0-kb *Bam*HI fragment was subcloned (pBS1007) (Fig. 5B), and its nucleotide (nt) sequence was determined.

10 **Analysis of the DNA sequences of the *sgcA* and *sgcB* genes.**

Two complete open reading frames (ORFs) (*sgcA* and *sgcB*) were identified within the 3.0-kb *Bam*HI fragment of pBS1007, the 3,035-nt sequence of which is shown in Figure 6. The *sgcA* gene most likely begins with an ATG at nt 101, preceded by a probable ribosome biding site (RBS), GGAGG, and ends with a TGA stop codon at nt 1099. SgcA

15 should therefore encode a 332-amino acid protein with a molecular weight of 36,341 and an isoelectric point of 6.01. A Gapped-BLAST search showed that the deduced *sgcA* gene product is highly homologous to various putative and known NGDH enzymes from antibiotic-producing actinomycetes, including Gdh from the erythromycin biosynthesis gene cluster in *Saccharopolyspora erythraea* (64% identity and 70% similarity) (Linton *et al.*

20 (1995) *Gene* 153: 33-40), MtmE from the mithramycin biosynthesis gene cluster in *Streptomyces argillaceus* (64% identity and 68% similarity) (Lombo *et al.* (1997) *J. Bacteriol.* 179: 3354-3357), and TylA2 from the tylosin biosynthesis gene cluster in *Streptomyces fradiae* (62% identity and 68% similarity) (Merson-Davies and Cundliffe (1994) *Mol. Microbiol.* 13: 349-355) (Fig. 7). A conserved sequence of 14 amino acid

25 residues close to the N-termini can be easily identified in these proteins, which has been described as a $\beta\alpha\beta$ fold with an NAD^+-binding motif, GxGxxG, (Fig. 7, boxed), consistent with their biochemical role in deoxyhexose biosynthesis (Liu and Thorson (1994) *Ann. Rev. Microbiol.* 48: 223-256; Piepersberg (1997) pp. 81-163. In *Biotechnology of antibiotics,* 2nd ed. W. R. Strohl (ed). Marcel Dekker, New York). The function of Gdh and MtmE as TDP-

30 glucose 4,6-dehydratases, requiring NAD^+ as a cofactor, has been confirmed by an enzyme assay following expression of the *gdh* (Linton *et al.* (1995) *Gene* 153: 33-40) and *mtmE* gene (Lombo *et al.* (1997) *J. Bacteriol.* 179: 3354-3357) in *E. coli,* respectively, and by

-46-

purification of the Gdh protein from *Sacc. erythraea* (Vara *et al.* (1989) *J. Bacteriol.* 171: 5872-5881). From these data, it is reasonable to suggest that *sgcA* encodes the NGDH enzyme required for the biosynthesis of the 4,6-dideoxy-4-dimethylamino-5-methylrhamnose moiety of the C-1027 chromophore.

5   Transcribed in the same direction as *sgcA*, the *sgcB* gene is located 43 nt downstream of *sgcA*. It should begin with a GTG at nt 1143, preceded by a probable RBS, AGGAG, and end with a TGA at nt 2708 (Fig. 6). Correspondingly, *sgcB* should therefore encode a 521-amino acid protein with a molecular weight of 52,952 and an isoelectric point of 4.64. Database comparison of the deduced *sgcB* product revealed that SgcB is closely

10 related to a family of membrane efflux pumps, such as LfrA from *Mycobacterium smegmatis* (43% identity and 50% similarity, protein accession number AAC43550) (Takiff *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93: 362-366), OrfA from *Streptomyces cinnamomeus* (42% identity and 47% similarity, protein accession number AAB71209) (Sommer *et al.* (1997) *Appl. Environ. Microbiol.* 63: 3553-3560), and RifP from the rifamycin biosynthesis

15 gene cluster in *Amycolatopsis mediterranei* (35% identity and 44% similarity, protein accession number AAC01725) Augus *et al.* (1998) *Chem. Biol.* 5: 69-79). These proteins are membrane-localized transporters involved in the transport of antibiotics (conferring resistance), sugars, and other substances. While direct evidence is lacking for RifP conferring rifamycin resistance in *A. mediterranei* by transporting it out of the cells (August

20 *et al.* (1998) *Chem. Biol.,* 5: 68-79), it has been proven that LfrA employs the transmembrane proton gradient in an antiporter mode to drive the efflux of intracellular antibiotics, resulting in fluoroquinolone resistance in *M. smegmatis* (Takiff *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93: 362-366). On the basis of the high degree of amino acid sequence conservation, an equivalent role could be proposed for SgcB, conferring resistance

25 by exporting C-1027 from *S. globisporus*.

The *cagA* gene is clustered with the *sgcA* and *sgcB* locus.

       To determine if *cagA* is clustered with the *sgcA* and *sgcB* locus, PCR primers were designed according to the flanking regions of *cagA* (Sakata *et al.* (1992) *Biosci. Biotech. Biochem.* 56: 1592-1595). A single product with the predicted size of 0.73 kb was

30 indeed amplified from several of the overlapping cosmids (which cover the 75-kb *sgc* cluster), including pBS1004 and pBS1005, the identity of which as *cagA* was confirmed by sequencing. Restriction enzyme mapping and Southern hybridization analysis localized

cagA to a single 4.0-kb *Bam*HI fragment that is approximately 14 kb upstream of the *sgcA,B* locus (Fig. 5B). The 4.0-kb *Bam*HI fragment was subcloned (pBS1008), and its nt sequence was determined, revealing the *cagA* gene along with two additional ORFs (data not shown) (Fig. 5). As reported earlier, *cagA* encodes a 142-amino acid protein that is processed by

5   cleavage of a 32-amino acid lead peptide to yield the mature CagA apoprotein (Sakata *et al.* (1992) *Biosci. Biotech. Biochem.* 56: 1592-1595).

Disruption of the *sgcA* gene in *S. globisporus*.

         To examine if the cloned *sgc* cluster encodes C-1027 biosynthesis, *sgcA* was insertionally disrupted by a single crossover homologous recombination event to generate C-

10   1027-nonproducing mutant strains (Fig. 8A). Two plasmids were used—pBS1012 (a pOJ260 derivative) and pBS1013 (a pKC1139 derivative), each of which contain a 0.75-kb internal fragment from *sgcA* (Table 3). After introduction of pBS1012 into *S. globisporus* either by PEG-mediated protoplast transformation or *E. coli-S. globisporus* conjugation, transformants or exconjugants that were resistant to apramycin were isolated in all cases.

15   Since pBS1012 is derived from the *Streptomyces* non-replicating plasmid of pOJ260, these isolates must have resulted from integration of pBS1012 into the *S. globisporus* chromosome by homologous recombination. Plasmid pBS1013 was similarly introduced into *S. globisporus*. However, since pBS1013 is derived from pKC1139 that carries the temperature-sensitive *Streptomyces* replication origin from pSG5 and can replicate normally

20   at 28°C (Muth *et al.* (1989) *Mol. Gen. Genet.* 219: 341-348), these isolates were subjected to incubation at the non-permissive temperature of 37°C to eliminate free plasmids from the host cells. As expected, normal growth stopped except for the recombinants that continue to grow at 37°C, indicative of integration of pBS1013 into *S. globisporus* by homologous recombination. The apramycin-resistant *S. globisporus* SB1001 and *S. globisporus* SB1002

25   strains were chosen as representatives of mutant strains with disrupted *sgcA* gene resulted from integration of pBS1012 and pBS1013, respectively.

         To confirm that targeted *sgcA* disruption has occurred by a single crossover homologous recombination event, Southern analysis of the DNA from the mutant strains was performed as exemplified for *S. globisporus* SB1001 with either pOJ260 or the 0.75-kb

30   *Sac*II/*Kpn*I internal fragment of *sgcA* from pBS1010 as a probe. As shown in Fig. 8B, a distinctive band of the predicted size of 6.3 kb was detected with the pOJ260 vector as a probe in all mutant strains (lanes 2, 3, and 4); this band was absent from the wild-type strain

-48-

(lane 1). Complementarily, when using the 0.75-kb *Sac*II/*Kpn*I internal fragment of *sgcA* as a probe (Fig. 8C), the 3.0-kb band in the wild-type strain (lane 1) was split into two fragments with the size of 6.3 kb and 1.0 kb in the mutant strains (lanes 2, 3, and 4), as would be expected for disruption of *sgcA* by a single crossover homologous recombination

5   event.

S. globisporus SB1001 and *S.* globisporus SB1002 are C-1027-nonproducing mutants.

No apparent difference in growth characteristics and morphologies between the wild-type *S. globisporus* and mutant *S. globisporus* SB1001 and *S. globisporus* SB1002

10   strains was observed. While C-1027 production in the wild-type *S. globisporus* strain could be detected on day 3, peaked on day 5, and continued for a few more days, as judged by assaying the antibacterial activity of the culture supernant against *M. luteus* (Hu *et al.* (1988) *J. Antibiot.* 41: 1575-1579), C-1027 production is completely abolished in the *sgcA* mutant strains *S. globisporus* SB1001 and *S. globisporus* SB1002 (Fig. 9A). The latter phenotype

15   was identical to that of the AF40, AF44, and AF67 mutants, C-1027-nonproducing *S. globisporus* strains that have been characterized previously (Fig. 9A and 9C) (Mao, *et al.* (1997) *Chinese J. Biotechnol.* 13: 195-199).

In vivo complementation of *S.* globisporus SB1001.

The ability of the wild-type *sgcA* gene to complement the disrupted *sgcA* gene

20   was tested in the *S. globisporus* SB1001 strain. The construction of pBS1015, in which the expression of *sgcA* is under the control of the constitutive *ermE** promoter, was described in Materials and Methods. Both the pBS1015 construct and the pWHM3 vector as a control were introduced by transformation into the *S. globisporus* SB1001 mutant strains. Culture supernants from each transformant were bioassayed against *M. luteus* for C-1027 production.

25   pBS1015 restored C-1027 production to *S. globisporus* SB1001 to the wild-type level; no C-1027 production was detected in the control in which pWHM3 was introduced into *S. globisporus* BS1001 (Fig. 9B and 9C). A significant reduction of C-1027 production was observed when *S. globisporus* SB1001(pBS1015) was cultured under identical conditions but without thiostrepton (Fig. 9B vs. 6C), indicative that pBS1015 may be unstable in *S.*

30   *globisporus* SB1001 in the absence of antibiotic selection pressure.

Expression of *sgcB* in *S. globisporus.*

The effect of *sgcB* on C-1027 production was tested in the wild-type *S. globisporus* strain. The construction of pBS1018, in which the expression of *sgcB* is under the control of the constitutive *ermE** promoter, was described in Materials and Methods.

5   pBS1018 and the pKC1139 vector as a control were each introduced by conjugation into *S. globisporus*. Culture supernatants from each exconjugant were harvested on days 3, 4, and 5, and assayed for C-1027 production by determining the antibacterial activity against *M. luteus*. While no apparent difference for C-1027 production was observed between the *S. globisporus* and *S. globisporus* (pKC1139) strains, a significant increase in C-1027

10   production (150±25%) was evident in the early stage of *S. globisporus* (pBS1018) fermentation (Fig. 9D, day 3). However, such effect on C-1027 production leveled off as the fermentation proceeded and became insignificant when the culture reached the late stationary phase of fermentation (Fig. 9D, day 4 and 5).

Discussion.

15   Our inability to clone the putative enediyne PKS gene by PCR, with degenerate primers designed according to the highly conserved amino acid sequences of either type I or type II PKSs, or by DNA hybridization, with homologous type I or type II PKS as probes (data not shown), was unexpected, since feeding experiments by incorporation of [1-^{13}C]- and [1,2-^{13}C]acetate into the enediyne cores of esperamicin (Lam

20   *et al.* (1993) *J. Am. Chem. Soc.* 115: 12340-12345), dynemicin (Tokiwa *et al.* (1992) *J. Am. Chem. Soc.* 114: 4107-4110), and neocarzinostatin (Hensens *et al.* (1989) *J. Am. Chem. Soc.* 111: 3295-3299) supported their polyketide origin. Although the enediyne cores are structurally distinct from either the reduced or aromatic polyketides, the biosynthesis of which is well characterized by type I or type II PKS, respectively, it could be imagined that

25   an enediyne PKS catalyzes the biosynthesis of a polyunsaturated linear heptaketide intermediate that is subsequently cyclized into the enediyne core structure (Hu *et al.* (1994) *Mol. Microbiol.* 14: 163-172; Spaink *et al.* (1991) *Nature* 354: 125-130; Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188). Alternatively, Hensens and co-workers proposed a fatty acid origin for the enediyne core that was also consistent with the isotope labeling

30   results. These authors suggested oleate as a precursor that is shortened by loss of carbons from both ends and is desaturated via the oleate-crepenynate pathway to furnish the enediyne core (Hensens *et al.* (1989) *J. Am. Chem. Soc.* 111: 3295-3299). The latter pathway

resembles polyacetylene biosynthesis in higher plants and fungi and requires an acetylene forming enzyme—a plant gene encoding such an enzyme was identified recently (Lee *et al.* (1998) *Science* 280: 915-918). Our DNA sequence analysis of approximately 60 kb of the *sgc* gene cluster, fails to reveal any gene that resembles PKS.

5   Although little is known about the resistance mechanism for the enediyne antibiotics in general, the apoproteins of the chromoprotein type of enediynes could be viewed as resistance elements that confer self-resistance to the producing organisms by drug sequestration (Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188). Such a resistance mechanism is in fact well established in antibiotic-producing actinomycetes, for example,

10   BlmA, the bleomycin-binding protein from *Streptomyces verticillus* (Shen *et al.* (1999) *Bioorg. Chem.* 27: 155-171). Given the fact that antibiotic production genes have invariably been found to be clustered in one region of the microbial chromosome, consisting of structural, resistance, and regulatory genes, we adopted a strategy to clone the *sgc* gene cluster by mapping a putative C-1027 structural gene to the previously cloned *cagA* gene,

15   considered as a resistance gene that encodes the C-1027 apoprotein.

        We chose NGDH as the putative C-1027 structural gene on the basis of the 4,6-dideoxy-4-dimethylamino-5-methylrhamnose moiety of the C-1027 chromophore. It has been well established that all deoxyhexoses could be derived from the common intermediate of 4-keto-6-deoxyglucose nucleoside diphosphate, the biosynthesis of which from glucose

20   nucleoside diphosphate is catalyzed by an NGDH enzyme. We cloned the NGDH gene from *S. globisporus* by PCR and used it as a probe to screen an *S. globisporus* genomic library, resulting in the isolation of the 75-kb *sgc* gene cluster. DNA sequence analysis of a 3.0-kb *Bam*HI fragment of the *sgc* cluster confirmed the presence of the NGDH protein, encoded by *sgcA*, along with *sgcB* that encodes a transmembrane efflux protein (Fig. 6). The *cagA* gene

25   indeed resides approximately 14 kb upstream of *sgcA* (Fig. 5); DNA sequence analysis of a 4.0-kb *Bam*HI fragment confirmed the identity of *cagA* along with two additional ORFs (data not shown). These results underline once again the effectiveness of cloning natural product biosynthesis gene clusters by exploiting the clustering phenomenon between resistance and structural genes.

30   The involvement of the cloned gene cluster in C-1027 biosynthesis was demonstrated by disrupting the *sgcA* gene to generate *S. globisporus* mutants, the ability of which to produce C-1027 was completely abolished (Fig. 9A), and by complementing the *sgcA* mutants in vivo upon expression of *sgcA* in trans to restore C-1027 production (Fig. 9B

and 6C). These data unambiguously establish that *sgcA* is essential for C-1027 production, and thus support the conclusion that the cloned gene cluster encodes C-1027 biosynthesis. It should be pointed out that, although the *sgcA* mutants *S. globisporus* SB1001 and *S. globisporus* SB1002 were characterized as C-1027-nonproducing on the basis of the

5   antibacterial assay alone (Fig. 9A), this phenotype was identical to that of the controls of the AF40, AF44, and AF67 mutants (Fig. 9A and 9C). The latter strains were isolated previously upon randomly mutagenizing the wild-type *S. globisporus* strain with acriflavine and confirmed to be C-1027-nonproducing by both the antibacterial bioassay and an antitumor spermatogonial assay (Mao, *et al.* (1997) *Chinese J. Biotechnol.* 13: 195-199),

10 providing strong support to the current study. Gene disruption and complementation in *S. globisporus* were made possible by the recently developed genetic system that allowed us to introduce plasmid DNA into *S. globisporus* via either PEG-mediated protoplast transformation (Hopwood *et al.* (1985) *Genetic manipulation of Streptomyces: a laboratory manual.* John Innes Foundation, Norwich, UK) or *E. coli-S. globisporus* conjugation

15 (Bierman *et al.* (1992) *Gene* 116: 43-69; Matsushima and Baltz (1996) *Microbiology* 142: 261-267; Matsushima *et al.* (1994) *Gene* 146: 39-45) for analyzing the *sgc* biosynthesis gene cluster *in vivo*. Given the difficulties encountered with calicheamicin biosynthesis in *Micromonospora echinospora*, into which all attempts to introduce plasmid DNA have failed (Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188), the latter results underscore the

20 importance of selecting C-1027 as a model system for enediyne biosynthesis so that many of the genetic tools developed in *Streptomyces* species can now be directly applied to the study of enediyne biosynthesis.

Finally, the function of *sgcB* was probed by examining C-1027 production, following expression of the gene in the wild-type *S. globisporus* strain. Database

25 comparison of the deduced amino acid sequence clearly suggested SgcB as a transmembrane efflux protein, conferring resistance by exporting C-1027 out of the cell. Hence, in addition to CagA, SgcB could be viewed as the second resistance element identified for C-1027 biosynthesis. Multiple resistance genes have been identified in numerous antibiotic biosynthesis gene clusters (Hopwood (1997) *Chem. Rev.* 97: 2465-2497). It could be

30 imagined that CagA and SgcB function cooperatively to provide resistance—the C-1027 chromophore is first sequestered by binding to the preaproprotein CagA to form a complex, which is then transported out of the cell by the efflux pump SgcB and processed by removing the leader peptide to yield the chromoprotein, although we do not have any experimental

data to substantiate this speculation. Since it is known that yields for antibiotic production could be profoundly altered by the introduction of extra copies of regulatory, resistance, or structural genes into wild-type organisms (Hutchinson (1994) *Bio/Technology* 12: 375-380), we tested the effect of overexpressing *sgcB* in *S. globisporus* on C-1027 production. While no apparent adverse effect on C-1027 production was observed upon introduction of the pKC1139 vector into *S. globisporus* (data not shown), a significant increase in C-1027 production (150±25%) was observed in the early stage of *S. globisporus* (pBS1017) fermentation (Fig. 9D, day 3), supporting the predicted function for SgcB in C-1027 biosynthesis. We propose that C-1027 resistance could be a limiting factor at the onset of C-1027 production, which is circumvented by the extra copy of the plasmid-born *sgcB*, and overexpression of *sgcB* under the control of the constitutive *ermE** promoter results in increase of C-1027 production. However, as the *S. globisporus* (pBS1017) fermentation proceeds to its stationary phase, C-1027 resistance is no longer a limiting factor for overall C-1027 production, and the effect of extra copy of SgcB on C-1027 production consequently became insignificant (Fig. 9D, day 5).

In conclusion, genetic analysis of enediyne biosynthesis has heretofore met with little success in spite of considerable effort (Thorson *et al.* (1999) *Bioorg. Chem.*, 27: 172-188). The localization of the *sgc* gene cluster and characterization of the *sgcA* and *sgcB* genes have now provided an excellent basis for genetic and biochemical investigations and/or modification of C-1027 biosynthesis, and gene disruption and overexpression in *S. globisporus* clearly demonstrated the potential to construct enediyne-overproducing strains and to produce novel enediynes that may have enhanced potency as novel anticancer drugs using combinatorial biosynthesis and targeted mutagenesis. We envisage that the results from C-1027 biosynthesis should facilitate the cloning and characterization of biosynthesis gene clusters of other enediyne antibiotics in *Streptomyces* as well as in other actinomycetes, and could have a great impact on the overall field of combinatorial biosynthesis.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

CLAIMS

What is claimed is:

1.   An isolated nucleic acid comprising a nucleic acid selected from the group consisting of

5   a nucleic acid encoding any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA);

a nucleic acid encoding a polypeptide encoded by any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA); and

a nucleic acid amplified by polymerase chain reaction (PCR) using

10   primer pairs that amplify any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA).

2.   The isolated nucleic acid of claim 1, wherein said nucleic comprises a nucleic acid encoding at least two open reading frames (ORFs) selected from the group consisting of ORF-1 through ORF 42, excluding ORF 9 (cagA).

15   3.   The isolated nucleic acid of claim 1, wherein said nucleic comprises a nucleic acid encoding at least three open reading frames (ORFs) selected from the group consisting of ORF-1 through ORF 42, excluding ORF 9 (cagA).

4.   An isolated nucleic acid comprising a nucleic acid that specifically hybridizes under stringent conditions to an open reading frame (ORF) of the C-1027

20   biosynthesis gene cluster, excluding ORF 9 (cagA), and can substitute for the ORF to which it specifically hybridizes to direct the synthesis of an enediyne.

5.   The isolated nucleic acid of claim 4, wherein said isolated nucleic acid comprises a nucleic acid that specifically hybridizes under stringent conditions to a nucleic acid selected from the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -

25   2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 8, ORF 10, ORF 11, ORF 12, ORF 13, and ORF 14.

6.   The isolated nucleic acid of claim 4, wherein said isolated nucleic acid comprises a nucleic acid that specifically hybridizes under stringent conditions to a nucleic

acid selected from the group consisting of ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42.

7.   The isolated nucleic acid of claim 5, wherein said isolated nucleic acid comprises a nucleic acid selected from the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 8, ORF 10, ORF 11, ORF 12, ORF 13, and ORF 14.

8.   The isolated nucleic acid of claim 6, wherein said isolated nucleic acid comprises a nucleic acid selected from the group consisting of ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42.

9.   The isolated nucleic acid of claim 4, wherein said nucleic acid comprises a nucleic acid that is a single nucleotide polymorphism (SNP) of a nucleic acid selected from the group consisting of ORF -7, ORF -6, ORF -5, ORF -4, ORF -3, ORF -2, ORF -1, ORF 0, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 81, ORF 1, ORF 2, ORF 3, ORF 4, ORF 5, ORF 6, ORF 7, ORF 8, ORF 10, ORF 11, ORF 12, ORF 13, ORF 14, ORF 15, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 23, ORF 24, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 34, ORF 35, ORF 36, ORF 37, ORF 38, ORF 39, ORF 40, ORF 41, and ORF 42.

10.   An isolated gene cluster comprising open reading frames encoding polypeptides sufficient to direct the assembly of a C-1027 enediyne or a C-1027 enediyne analogue.

11.   The gene cluster of claim 10, wherein said gene cluster is present in a bacterium.

12.   The gene cluster of claim 11, wherein said gene cluster is present in a bacterium selected from the group consisting of *Actinomycetes, Actinoplanetes, Actinomadura, Micromonospora, and Streptomycetes.*

-56-

13.   The gene cluster of claim 11, wherein said gene cluster is present in a bacterium selected from the group consisting *Streptomyces globisporus*, *Streptomyces lividans*, *Streptomyces coelicolor*, *Micromonospora echinospora* spp. *calichenisis*, *Actinomadura verrucosopora*, *Micromonospora chersina*, *Streptomyces carzinostaticus*, and *Actinomycete* L585-6.

14.   The gene cluster of claim 13, wherein one or more open reading frames is operatively linked to a heterologous promoter.

15.   An isolated polypeptide comprising a catalytic domain encoded by a nucleic acid of a C-1027 gene cluster wherein said nucleic acid comprises a nucleic acid selected from the group consisting of

a nucleic acid encoding any of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA); and

a nucleic acid amplified by polymerase chain reaction (PCR) using any one of the primer pairs identified in Tables I and II that specifically amplify one or more of (ORFs) -7 through 42, excluding ORF 9 (cagA).

16.   The polypeptide of claim 15, wherein said polypeptide is encoded by at least two open reading frames selected from the group consisting of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA).

17.   The polypeptide of claim 15, wherein said polypeptide is encoded by at least three open reading frames selected from the group consisting of C-1027 open reading frames (ORFs) -7 through 42, excluding ORF 9 (cagA).

18.   An expression vector comprising a nucleic acid of any one of claims 1 through 9.

19.   A host cell transformed with an expression vector of claim 18.

20.   The host cell of claim 19, wherein said cell is transformed with an exogenous nucleic acid comprising a gene cluster encoding polypeptides sufficient to direct the assembly of a C-1027 enediyne or a C-1027 enediyne analogue.

21.   The host cell of claim 19, wherein said host cell is a bacterium.

-57-

22.   The host cell of claim 21, wherein said bacterium is selected from the group consisting of Actinomycetes, *Actinoplanetes, Actinomadura, Micromonospora,* and *Streptomycetes.*

23.   The host cell of claim 21, wherein said bacterium is selected from the group consisting of *Streptomyces globisporus, Streptomyces lividans, Streptomyces coelicolor, Micromonospora echinospora* spp. *calichenisis, Actinomadura verrucosopora, Micromonospora chersina, Streptomyces carzinostaticus, and Actinomycete* L585-6.

24.   A method of chemically modifying a biological molecule, said method comprising contacting a biological molecule that is a substrate for a polypeptide encoded by a C-1027 biosynthesis gene cluster open reading frame, with a polypeptide encoded by a C-1027 biosynthesis gene cluster open reading frame whereby said polypeptide chemically modifies said biological molecule.

25.   The method of claim 24, wherein said polypeptide is an enzyme selected from the group consisting of a hydroxylase, a homocysteine synthase, a dNDP-glucose dehydrogenase, a citrate carrier protein, a C-methyl transferase, an N-methyl transferase, an aminotransferase, a CagA apoprotein, an NDP-glucose synthase, an epimerase, an acyl transferase, a coenzyme F390 synthase, and epoxidase hydrolase, an anthranilate synthase, a glycosyl transferase, a monooxygenase, a type II condensation protein, an aminomutase, a type II adenylation protein, an O-methyl transferase, a P-450 hydroxylase, an oxidoreductase, and a proline oxidase.

26.   The method of claim 24, wherein said method comprising contacting said biological molecule with at least two different polypeptides encoded by C-1027 biosynthesis gene cluster open reading frames.

27.   The method of claim 24, wherein said method comprising contacting said biological molecule with at least three different polypeptides encoded by C-1027 biosynthesis gene cluster open reading frames.

28.   The method of claim 24, wherein said contacting is in a host cell.

29.   The method of claim 28, wherein said host cell is a bacterium.

-58-

30.   The method of claim 24, wherein said contacting *ex vivo*.

31.   The method of claim 28, wherein said biological molecule is an endogenous metabolite produced by said host cell.

32.   The method of claim 28, wherein said biological molecule is an exogenous supplied metabolite.

33.   The method of claim 28, wherein said host cell is a eukaryotic cell.

34.   The method of claim 33, wherein said eukaryotic cell is selected from the group consisting of a mammalian cell, a yeast cell, a plant cell, a fungal cell, and an insect cell.

35.   The method of claim 28, wherein said host cell synthesizes sugars and glycosylates the biological molecule.

36.   The method of claim 35, wherein said host cell synthesizes deoxysugars.

37.   The method of claim 24, wherein said method further comprises contacting said biological molecule with a polyketide synthase or a non-ribosomal polypeptide synthetase.

38.   The method of claim of claim 24, wherein said contacting is in a bacterial cell.

39.   The method of claim of claim 24, wherein said contacting is *ex vivo*.

40.   The method of claim 24, wherein said method comprises contacting said biological molecule with at substantially all of the polypeptides encoded by C-1027 biosynthesis gene cluster open reading frames and said method produces an enediyne or enediyne analogue.

41.   The method of claim 24, wherein said biological molecule is a fatty acid and said biological molecule is contacted with a C-1027 orf polyeptide selected from the

group consisting of an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase.

42.   The method of claim 41, wherein said biological molecule is a fatty acid and said biological molecule is contacted with a plurality of C-1027 orf polypeptides

5   comprising an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase.

43.   The method of claim 42, wherein said biological molecule is contacted with polypeptides encoded by ORF17, ORF20, ORF21, ORF29, ORF30, ORF32, ORF35, and ORF38.

10   44.   The method of claim 41, wherein said biological molecule is contacted with polypeptides encoded by ORF 15, ORF 16, ORF 28, ORF3, ORF 14, and ORF 13.

45.   The method of claim 44 wherein said biological molecule is also contacted with polypeptides encoded by ORF 4 and ORF 3.

46.   The method of claim 24, wherein said method comprises contacting a

15   sugar with one or more C-1027 open reading frame polypeptides selected from the group consisting of a dNDP-glucose synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase.

47.   The method of claim 46, wherein said method comprises contacting a dNDP-glucose with a plurality of C-1027 open reading frame polypeptides comprising a

20   dNDP-glucose synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase.

48.   The method of claim 24, wherein said method comprises contacting an amino acid with one or one or more C-1027 open reading frame polypeptides selected from the group consisting of a hydroxylase, an aminomutase, a type II NRPS condensation

25   enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein.

49.   The method of claim 48, wherein said method comprises contacting an amino acid with a plurality of C-1027 open reading frame polypeptides comprising a

hydroxylase, a halogenase, an aminomutase, a type II NRPS condensation enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein.

50. The method of claim 48, wherein said amino acid is a tyrosine.

51. A method of synthesizing a chromaprotein type enediyne core, said method comprising contacting a fatty acid with one or more C-1027 orf polypeptides selected from the group consisting of an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase.

52. The method of claim 51, wherein said fatty acid is contacted with a plurality of C-1027 orf polypeptides comprising an epoxide hydrase, a monooxygenase, an iron-sulfer flavoprotein, a p-450 hydroxylase, an oxidoreductase, and a proline oxidase.

53. The method of claim 52, wherein said fatty acid is contacted with polypeptides encoded by ORF17, ORF20, ORF21, ORF29, ORF30, ORF32, ORF35, and ORF38.

54. A method of synthesizing a deoxysugar, said method comprising contacting a sugar with one or more C-1027 open reading frame polypeptides selected from the group consisting of a dNDP-glucose synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase.

55. The method of claim 54, wherein said method comprises contacting a dNDP-glucose with a plurality of C-1027 open reading frame polypeptides comprising a dNDP-glucose synthase, a dNDP glucose dehydratase, an epimerase, an aminotransferase, a C-methyltransferase, an N-methyltransferase, and a glycosyl transferase.

56. The method of claim 55, wherein said dNDP-glucose is contacted with polypeptides encoded by ORF17, ORF20, ORF21, ORF29, ORF30, ORF32, ORF35, and ORF38.

57. A method of synthesizing a beta amino acid, said method comprising contacting an amino acid with one or one or more C-1027 open reading frame polypeptides selected from the group consisting of a hydroxylase, an aminomutase, a type II NRPS

condensation enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein.

58.   The method of claim 57, wherein said method comprises contacting an amino acid with a plurality of C-1027 open reading frame polypeptides comprising a hydroxylase, a halogenase, an aminomutase, a type II NRPS condensation enzyme, a type II NRPS adenylation enzyme, and a type II peptidyl carrier protein.

59.   The method of claim wherein said amino acid is contacted with polypeptides encoded by ORF 4, ORF11, ORF24, ORF23, ORF25, and ORF26.

60.   The method of claim 57, wherein said amino acid is a tyrosine.

61.   A method of synthesizing an enediyne or an enediyne analogue said method comprising:

culturing a cell comprising a recombinantly modified C-1027 gene cluster under conditions whereby said cell expresses said enediyne or enediyne analogue; and

recovering said enediyne or enediyne analogue.

62.   The method of claim 61, wherein said gene cluster is present in a bacterium.

63.   The gene cluster of claim 62, wherein said gene cluster is present in a bacterium selected from the group consisting of Actinomycetes, *Actinoplanetes, Actinomadura, Micromonospora,* and *Streptomycetes.*

64.   The gene cluster of claim 62, wherein said gene cluster is present in a bacterium selected from the group consisting *Streptomyces globisporus, Streptomyces lividans, Streptomyces coelicolor, Micromonospora echinospora* spp. *calichenisis, Actinomadura verrucosopora, Micromonospora chersina, Streptomyces carzinostaticus, and Actinomycete* L585-6.

65.   The method of claim 61, wherein said gene cluster is present in a eukaryotic cell.

66. The method of claim 65, wherein said eukaryotic cell is selected from the group consisting of a mammalian cell, a yeast cell, a plant cell, a fungal cell, and an insect cell.

67. The method of claim 61, wherein said host cell synthesizes sugars and glycosylates said enediyne or enediyne analogue.

68. The method of claim 67, wherein said host cell synthesizes deoxysugars.

69. A method of making a cell resistant to an enediyne or an enediyne metabolite, said method comprising expressing in said cell one or more isolated C-1027 open reading frame nucleic acids that encode a protein selected from the group consisting of a CagA apoprotein, a SgcB transmembrane efflux protein, a transmembrane transport protein, a Na+/H+ transporter, an ABC transport, a glycerol phosphate tranporter, and a UvrA-like protein.

70. The method of claim 69, wherein said isolated C-1027 open reading frame nucleic acids are selected from the group consisting of ORF 9, ORF2, ORF 27, ORF 0, ORF 1 c-terminus, ORF 2, and ORF 1 N-terminus.

71. The method of claim 69, wherein said cell is a bacterial cell.

GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR ANTIBIOTIC C-1027

ABSTRACT OF THE DISCLOSURE

5   This invention provides nucleic acid sequences and characterization of the gene cluster responsible for the biosynthesis of the enediyne C-1027 (produced by *Streptomyces globisporus*). Methods are provided for the biosynthesis of enediynes, enediyne analogs and other biological molecules.

10

15

20

FILE: c:_docs\2500 uc ott\128us1\2500.128wo0 enediyne.ap1.doc

C-1027 chromophore   **Benzenoid diradical**   **DNA cleavage**

Fig. 1

R = enediyne core
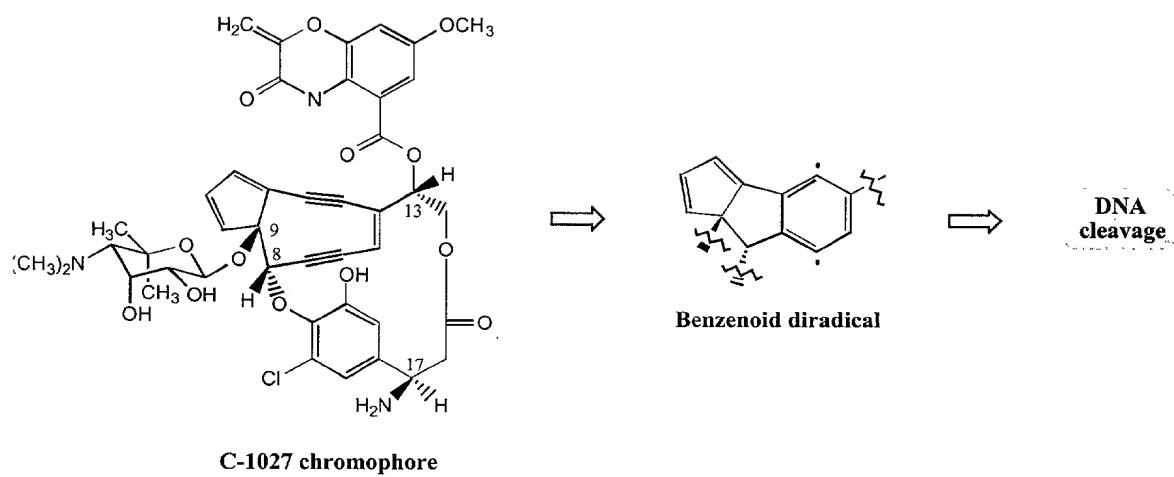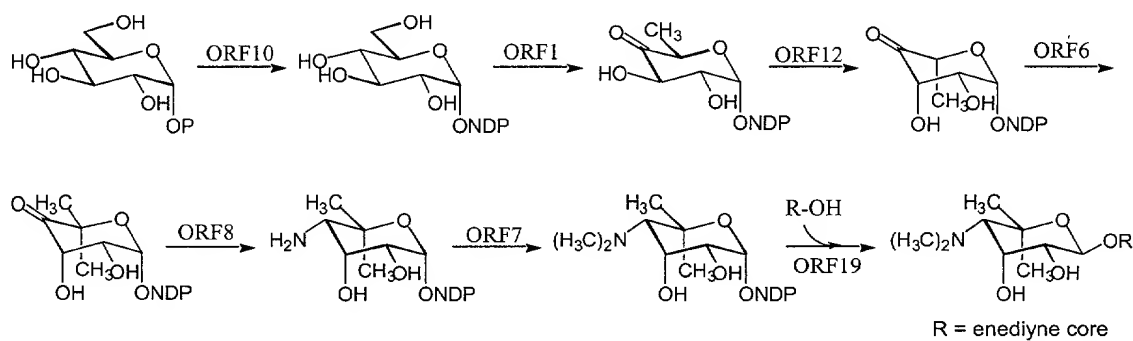
ORF10: dNDP-glucose synthase, 355 aa   ORF6: C-methyltransferase, 423 aa
ORF1: dNDP-glucose dehydratase, 332 aa   ORF7: N-methyltransferase, 244 aa
ORF12: epimerase, 192 aa   ORF19: glycosyl transferase, 459 aa
ORF8: aminotransferase, 410 aa

Fig. 2
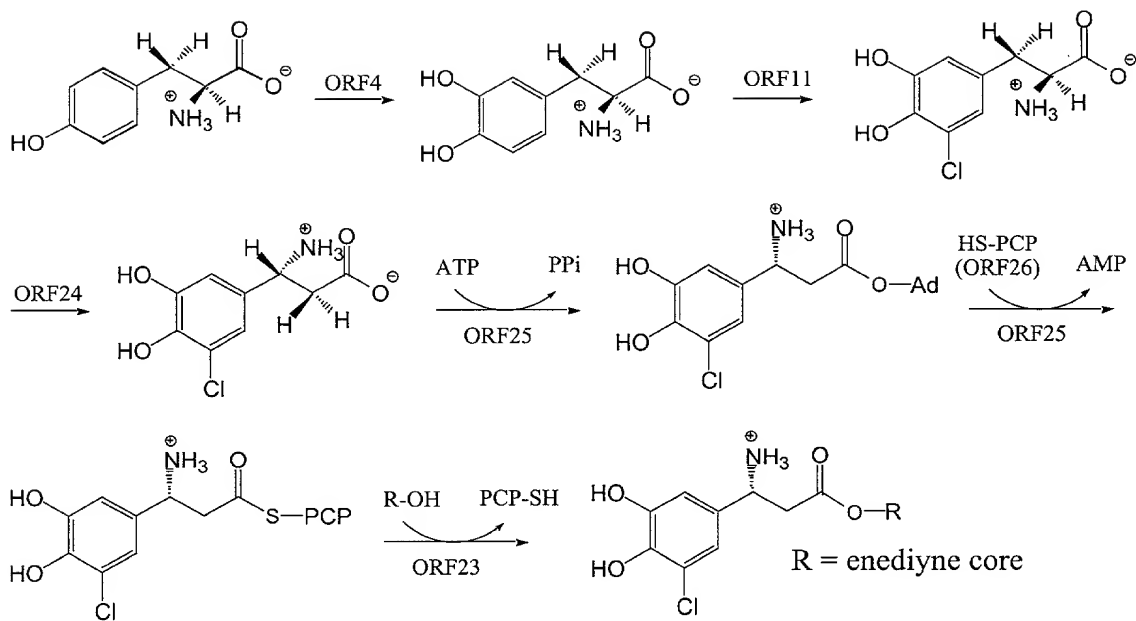
ORF4: Hydroxylase, 527 aa   ORF23: Type II NRPS condensation enzyme, 459 aa
ORF11: Hydroxylase/halogenase, 492/494 aa ORF25: Type II NRPS adenylation enzyme, 716 aa
ORF24: Aminomutase, 539 aa   ORF26: Type II peptidyl carrier protein, 93 aa

Fig. 3A

-3-
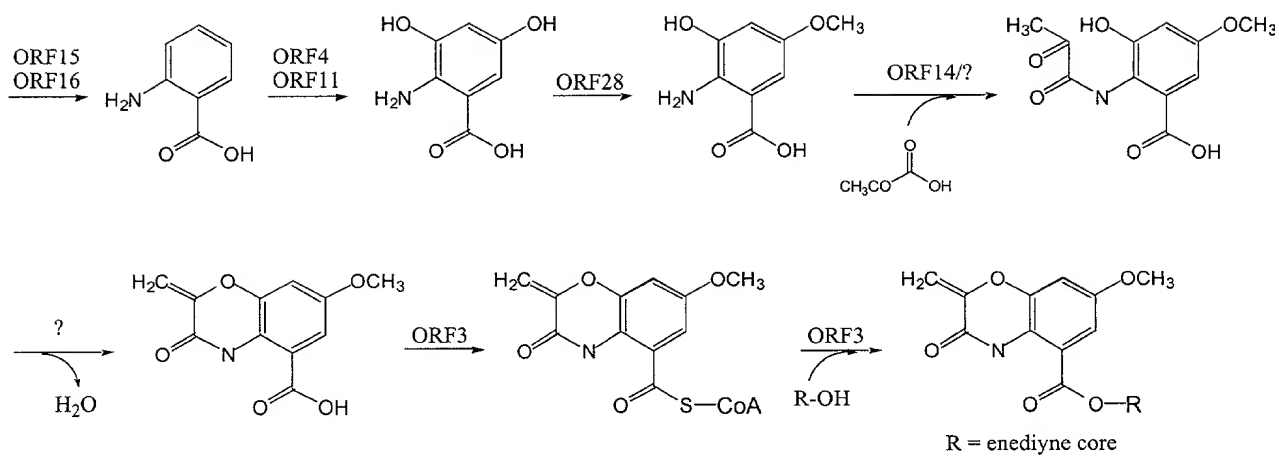
Fig. 3B

ORF15: Anthranilate synthase I, 493 aa   ORF3: Coenzyme F390 synthetase, 463 aa
ORF16: Anthranilate synthase II, 220 aa   ORF14: Coenzyme F390 synthetase, 484 aa
ORF28: O-methyltransferase, 350 aa   ORF13: O-acyltransferase, 378 aa

Fatty acid

ORF17: Epoxide hydrase
ORF20: Monooxygenase
ORF21: Iron-sulfur flavoprotein
ORF29: P-450 hydroxylase
ORF30: Oxidoreductase
ORF32: Oxidoreductase
ORF35: Proline oxidase
ORF38: P-450 hydroxylase

ORF13: O-acyltransferase, ORF19:Glycosyl transferase, ORF23: Type II NRPS condensation enzyme

Fig. 4

-5-

Fig. 5A

Fig. 5B

Fig. 5C

Fig. 6

```
       BamHI    EcoRI
   1   GGATCCGGGAAGACCGAATTCCGCCCCCCAGCCCGGTCGAACTCGTATCGCTCCTGGTAGAACTGACGAAGCGTCATCGCC|TGAC|AAGGAGGCGGACCG   100
                                                                                          sgcA >

 101   ATGAGGATGCTGGTGACGGGCGGAGCCGGGTTTCATCGGCTCGCAGTTCGTCGCCACACTGCACGGCGAGCTGCCGGGTTCCGAGGACGCCGGGTGA   200
        M  R  M  L  V  T  G  G  A  G  F  I  G  S  Q  F  V  R  A  T  L  H  G  E  L  P  G  S  E  D  A  R  V  T

 201   CGGTCCTGGACAAGCTGACGTACTCCGGCAATCCGGCCAATCTCACCTCCGTGGCCGCCCATCCGCGCTACACCTTCGTCCAGGGCGACACCGTCGACCC   300
        V  L  D  K  L  T  Y  S  G  N  P  A  N  L  T  S  V  A  A  H  P  R  Y  T  F  V  Q  G  D  T  V  D  P

 301   GCGCGTCGTCGACGAGGTGGTCGCCGGACACGACGTCATCGTGCACTTCGCCGCCGAGTCGCACGTCGACCGCTCCATCGATACCGCTACGCGTTTCGTC   400
        R  V  V  D  E  V  V  A  G  H  D  V  I  V  H  F  A  A  E  S  H  V  D  R  S  I  D  T  A  T  R  F  V

 401   ACGACCAACGTGCTCGGGACCCAGACGCTGCTGGAAGCGGCTCTCCGGCACGGGGTCGGCCGGTTCGTGCACGTGTCGACCGACGAGGTCTACGGGTCGA   500
        T  T  N  V  L  G  T  Q  T  L  L  E  A  A  L  R  H  G  V  G  R  F  V  H  V  S  T  D  E  V  Y  G  S  I

 501   TCGCCTCCGGCTCATGGACCGAGGACACCCCGCTCGCCCCCAACGTCCCGTACGCGGCGTCGAAGGCGGGTTCGGACCTGATGGCGCTCGCCTGGCACCG   600
        A  S  G  S  W  T  E  D  T  P  L  A  P  N  V  P  Y  A  A  S  K  A  G  S  D  L  M  A  L  A  W  H  R

 601   CACCCGGGGCCTGGACGTCGTCGTCACCCGTTGCACGAACAACTACGGTCCCTACCAGTACCCCGAGAAGGTGATCCCGCTCTTCGTCACCAACATCCTC   700
        T  R  G  L  D  V  V  T  R  C  T  N  N  Y  G  P  Y  Q  Y  P  E  K  V  I  P  L  F  V  T  N  I  L

                                                        SacII
 701   GACGGGCTTGCGGGTGCCCCGTACGGGGACGGCGCGCACCGGCGCGACTGGCTGCACGTGTCCGACCACTGCCGGGCCATCCAGATGGTCATGAACTCCG   800
        D  G  L  R  V  P  L  Y  G  D  G  A  H  R  R  D  W  L  H  V  S  D  H  C  R  A  I  Q  M  V  M  N  S  G

 801   GCCGGGGCCGGGGAGGTCTACCACATCGGCGGCGGCACCGAACTCTCCAACGAGGAACTGACCGGGCTGCTCCTCACGGCGTGCGGCACTGACTGGTCCTG   900
        R  A  G  E  V  Y  H  I  G  G  G  T  E  L  S  N  E  E  L  T  G  L  L  L  T  A  C  G  T  D  W  S  C

 901   CGTGGACCGGGTGGCCGACCGGCAGGGGCACGACGACCGGCGCTACTCGCTGGACATCACGAAGATCCGGCAGGAACTGGGCTACGAGCCCCTGGTCGCCTTC  1000
        V  D  R  V  A  D  R  Q  G  H  D  D  R  R  Y  S  L  D  I  T  K  I  R  Q  E  L  G  Y  E  P  L  V  A  F

1001   GAGGACGGGCTGGCCGCGACGGTGAAGTGGTACCACGAGAACCGTTCGTGGTGGCAGCCGCTCAAGGAAGCCGCCGGGCTGCTCGACGCCGTCGGCTGAC  1100
        E  D  G  L  A  A  T  V  K  W  Y  H  E  N  R  S  W  W  Q  P  L  K  E  A  A  G  L  L  D  A  V  G  *

                   KpnI
1101   GGCCAGCCCACCGTGAACACCCCAGG|AAAG|GAGCCACCTCCGTGACAGCAGTCAAGGAGCCGACGTCCCGCGCAGGAGCGCGGAGTGGATCGCTCTCG  1200
                                    sgcB > M  T  A  V  K  E  P  T  S  R  A  G  R  R  E  W  I  A  L  V

1201   TCGTCCTCCTTGCCCACGATGCTGTTGATGCTGGACATCAACGTCCTCATGCTCGCTCTGCCGCAGTTGAGCGAGGATCTCGGCGCCAGCAGCACGCA  1300
        V  L  S  L  P  T  M  L  L  M  L  D  I  N  V  L  M  L  A  L  P  Q  L  S  E  D  L  G  A  S  S  T  Q

1301   ACAGCTGTGGATCACCGACATCTACGGATTCGCGATCGCCGGCTTCCTCGTGACCATGGGCACCCTCGGCGACCGGATCGGCCGCCGCCGTCTGCTC  1400
        Q  L  W  I  T  D  I  Y  G  F  A  I  A  G  F  L  V  T  M  G  T  L  G  D  R  I  G  R  R  R  L  L  L

1401   GGGGGCGGCGCCGTCTTCGCGGTCGTGTCGGTCGTCGCCTTCTCCGACAGCGCGGCGATGCTCGTCGTCAGCCGCGCCGTGCTCGGCGTCGCCGGGG  1500
        G  G  A  A  V  F  A  V  V  S  V  V  A  A  F  S  D  S  A  A  M  L  V  V  S  R  A  V  L  G  V  A  G  A

1501   CCACGGGTGATGCCGTCGCTCGCGCTCATCAGCAACATGTTCGAGGACCCCAAGGAGCGGGGCACCGGCGATCGCCATCGCCATGTGGGCGAGCGCCATGATGGCC  1600
        T  V  M  P  S  T  L  A  L  I  S  N  M  F  E  D  P  K  E  R  G  T  A  I  A  M  W  A  S  A  M  M  A

1601   CGGAGTCGCCCTCGGCCCGGCCGTCGGCGGCGGGCTCGTGCTCGCCGCCTTCTGGTGGGGATCGGTGTTCCTCATCGCCGTTCCGGTGATGCTGCTGGTG  1700
        G  V  A  L  G  P  A  V  G  G  G  L  V  L  A  A  F  W  W  G  S  V  F  L  I  A  V  P  V  M  L  L  V
```

Fig. 6 cont'd.

```
1701  GTCACCGGCCCCGTGCTGCTCACCGAGTCCCGCGACCCCGACGCGGCCGGACGGCTGGACCTGCTGCTGAGCGCGGGGCTCTCCCTCGCGACCGTGCTGCCGGTGA   1800
      V  T  G  P  V  L  L  T  E  S  R  D  P  D  A  G  R  L  D  L  L  S  A  G  L  S  L  A  T  V  L  P  V  I

1801  TCTACGGACTGAAGGAGCTGGCCCGGACCGGGTGGGACCCGCTCGCCGCGGGTGCCGTGGTCCTCGGCGTGATCTTCGGCGCGCTGTTCGTCCAGCGCCA   1900
      Y  G  L  K  E  L  A  R  T  G  W  D  P  L  A  A  G  A  V  V  L  G  V  I  F  G  A  L  F  V  Q  R  Q

1901  GCGGGCGGTTGGCGACCCCATGCTGGACCTCGGCCTCTTCGCCGACCGGACCCTGCGGGCGGGTCTGACGGTCAGTCTGGTCAACGCCGTCATCATGGGC   2000
      R  R  L  A  D  P  M  L  D  L  G  L  F  A  D  R  T  L  R  A  G  L  T  V  S  L  V  N  A  V  I  M  G

2001  GGGACCGGACTGATGGTCGCCCTGTACCTCCAGACGATCGCCGGGCTGCACTCCCCGTTGGCCGCGGCTGGTCTGTGGCTGCTGATCCCGGCCTGCATGCTCGTCG   2100
      G  T  G  L  M  V  A  L  Y  L  Q  T  I  A  G  H  S  P  L  A  A  A  G  L  W  L  L  I  P  A  C  M  L  V  V
                                                                                          SphI

2101  TGGGGCGTACAGCTGTCGAACCTGCTGGCCCAGCGGATGCCCCCTTCCCGGGTGCTCTCGGGGGACTGCTGATCGCCGCGGTCGGTCTGATCAC   2200
      G  V  Q  L  S  N  L  L  A  Q  R  M  P  P  S  R  V  L  L  G  G  L  L  I  A  A  V  G  Q  L  L  I  T

2201  CCAGTGGACACCGAGGAGACGGCCCTCCTCATCGCCGCTCTCTGTTCCGGCGGCGAGTTCGGAGTTCGGAGGCCTCGGGCATCGGGGCCGATCATG   2300
      Q  V  D  T  E  D  T  A  L  L  I  A  A  T  T  L  I  Y  F  G  A  S  P  V  G  P  I  T  T  G  A  I  M

2301  GGAGCCGGCCCCCGGAGAAGGCGGGGTGCCGCTCGTGTCCGGCGAGTTCGGAGTGGCCGCTCGGGGAGTCTGG   2400
      G  A  A  P  P  E  K  A  G  A  A  A  S  S  L  S  A  T  G  G  E  F  G  V  A  L  G  I  A  G  L  G  S  L  G

2401  GCACCGTCGTGTACAGCGGGGTCGAGGTGCCGGACGCGGCGGGCCCGGCCGACGCAGACGCCGCCCTGCACACGGC   2500
      T  V  V  Y  S  A  G  V  E  V  P  D  A  A  G  P  A  D  A  D  A  A  Q  E  S  I  A  G  A  L  H  T  A

2501  CGGTCAGCTGGCCACCGGGCAGCGCCCGACTCCGCTGGGACTCCGGCGGCGGGCCGGTGTTC   2600
      G  Q  L  A  P  G  S  A  D  A  L  L  D  S  A  R  A  A  F  T  S  G  V  Q  S  V  A  A  V  C  A  V  F

2601  TCCCTGGCGCTCGCCGTCCTCATCGGCACCCGGCTGCGCGACATTCCGGCGACCACGGCCACGGCTCAACCGG   2700
      S  L  A  L  A  V  L  I  G  T  R  L  R  D  I  S  A  M  D  H  G  H  G  E  E  P  A  E  N  D  A  Q  P  A

2701  CCACATGAGGCGCACTTCCGGAGATGCAACGGCCGCGTCGAGGTATGAGGATCACCTTCCGGGGTGCACCTGCACCGGAGTGGGAGTACT   2800
      T  *
        BamHI

2801  GGAACAGCACGGCGGAGACCATGCCCCGCCAGGAACTCGAACAGTGGAAGTGGCCGCAGGCTTCAGGCCGCCATGGACCAGCCGCCAGAAGGCTTTCGCCCTT   2900

2901  CTGGCGGGGAACGACTCCCCGAGAACATCACCTCCATGGCGGAACTACGCGGCGGGGTGCCTCGCCTCCTCCTGCGCCAAGGCCGACCTCCTGCGCCGGCGAAGCCGCG   3000
                                                                                            SacII

3001  TCTCCCCCTTACGGCCACCTGGCCCCTCGCTGCTGGATCC   3035
```

Fig. 7

```
Gdh        1:~~~MRVLVTGGAGFIGSHYVRQLLGGAYPAFAGADVVVLDKLTYAGNEENLRPVADDPRF: 57
TylA2      1:~~~MRVLVTGGAGFIGSHFTGQLLTGAYPDLGATRTVVLDKLTYAGNPANLEHVAGHPDL: 57
SgcA       1:~~~MRMLVTGGAGFIGSQFVRATLHGELPGSEDARVTVLDKLTYSGNPANLTSVAAHPRY: 57
MtmE       1:MTTTSILVTGGAGFIGSHYVRTLLGPR..GVPDVTVTVLDKLTYAGTLTNLAEVSDSDRF: 58
consensus  1:   mrvLVTGGAGFIGShyvr lL g  pa      v VLDKLTYaGn  NL  Va   prf: 60


Gdh       58:RFVRGDICEWDVVSEVMREVDVVVHFAAETHVDRSILGASDFVVTNVVGTNTLLQGALAA:117
TylA2     58:EFVRGDIADHGWWRRLMEGVGLVVHFAAESHVDRSIESSEAFVRTNVEGTRVLLQAAVDA:117
SgcA      58:TFVQGDTVDPRVVDEVVAGHDVIVHFAAESHVDRSIDTATRFVTTNVLGTQTLLEAALRH:117
MtmE      59:RFVRGDICDAPLVDDLLAVHDQVVVHFAAESHVDRSILGAADFVRTNVTGTQTLLDAALRQ:118
consensus 61: FVrGDi d  vv evm   dvvVHFAAEsHVDRSI  a  FV TNV GTntLL aAl  :120


Gdh      118:NVSKFVHVSTDEVYGTIEHGSWPEDHLLEPNSPYSAAKAGSDLIARAYHRTHGLPVCITR:177
TylA2    118:GVGRFVHISTDEVYGSIAEGSWPEDHPVAPNSPYAATKAASDLLALAYHRTYGLDVRVTR:177
SgcA     118:GVGRFVHVSTDEVYGSIASGSWTEDTPLAPNVPYAASKAGSDLMALAWHRTRGLDVVVTR:177
MtmE     119:GIETFVHISTDEVYGSIDAGSWPETAPVSPNSLYSAAKASSDLVALAYHRTHGLDVRVTR:178
consensus121:gv kFVHvSTDEVYGsI  GSWpEd pl PNspY A KAgSDLiAlAyHRThGLdV vTR:180


Gdh      178:CSNNYGPYQFPEKVLPLFITNLMDGRRVPLYGDGLNVRDWLHVTDHCRGIQLVAESGRAG:237
TylA2    178:CSNNYGPRQYPEKAVPLFTTNLLDGLPVPLYGDGGNTREWLHVDDHCRGVALVGAGGRPG:237
SgcA     178:CTNNYGPYQYPEKVIPLFVTNILDGLRVPLYGDGAHRRDWLHVSDHCRAIQMVMNSGRAG:237
MtmE     179:CSNNYGSHQFPEKVIPLFVTSLLDGREVPLYGDGTNVRDWLHVDDHVRAIELVRTGGRAG:238
consensus181:CsNNYGp QfPEKvlPLFiTnllDG  VPLYGDG n RdWLHV DHcRgi lV   GRaG:240


Gdh      238:EIYNIGGGTELTNKELTERVLELMGQDWSMVQPVTDRKGHDRRYSVDHTKISEELGYEPV:297
TylA2    238:VIYNIGGGTELTNAELTDRILELCGADRSALRRVADRPGHDRRYSVDTTKIREELGYAPR:297
SgcA     238:EVYHIGGGTELSNEELTQLLLDACGAGWDRVRYVTDRKGHDRRYSLDITKIRQELGYEPL:297
MtmE     239:EVYNIGGGTELSNKELTQLLLDACGAGWDRVRYVTDRKGHDRRYSVDCTKIRRELGYRPA:298
consensus241:eiYnIGGGTELtN ELT  vLe cG dws v  V DR GHDRRYSvD TKIr ELGY P :300


Gdh      298:VPFERGLAETIEWYRDNRAWWEPLKSAPDGGK~~~~:329
TylA2    298:TGITEGLAGTVAWYRDNRAWWEPLKRSPGGRELERA:333
SgcA     298:VAFEDGLAATVKWYHENRSWWQPLKEAAGLLDAVG~:332
MtmE     299:REFGDALAETVAWYRHHRAWWEPLTRAYGAVAA~~~:331
consensus301:  f egLA Tv WYrdnRaWWePLk a gg      :336
```

-9-

Fig. 8A
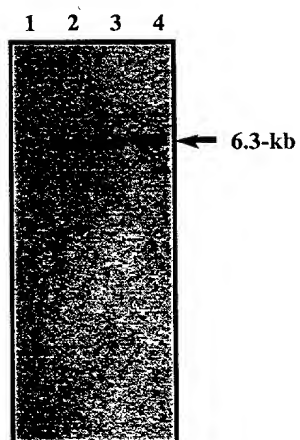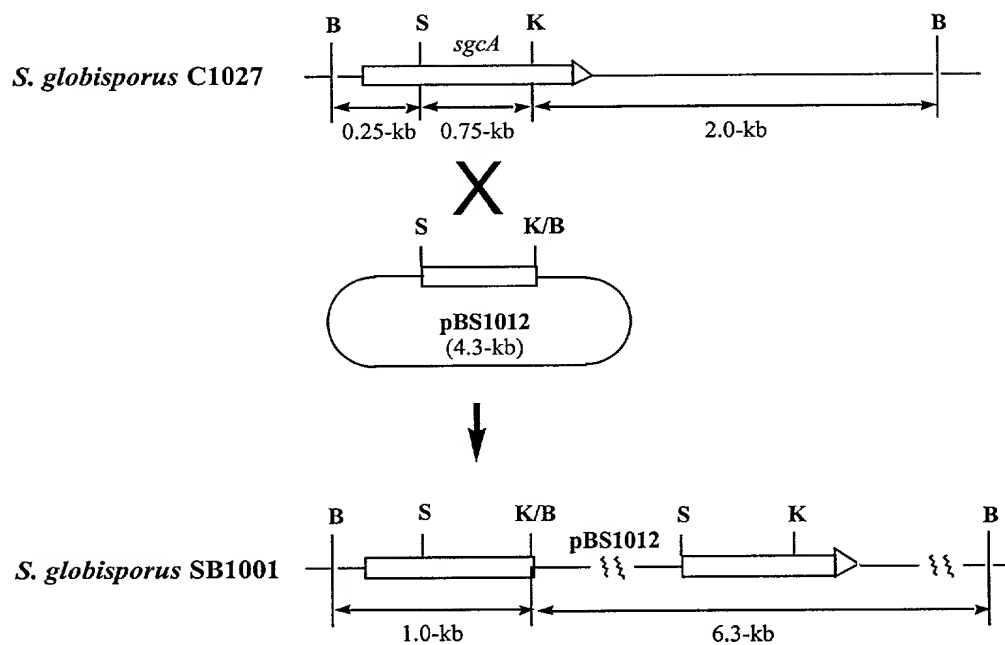
Fig. 8B
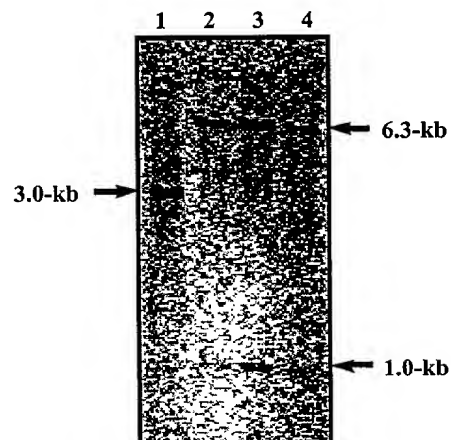
Fig. 8C
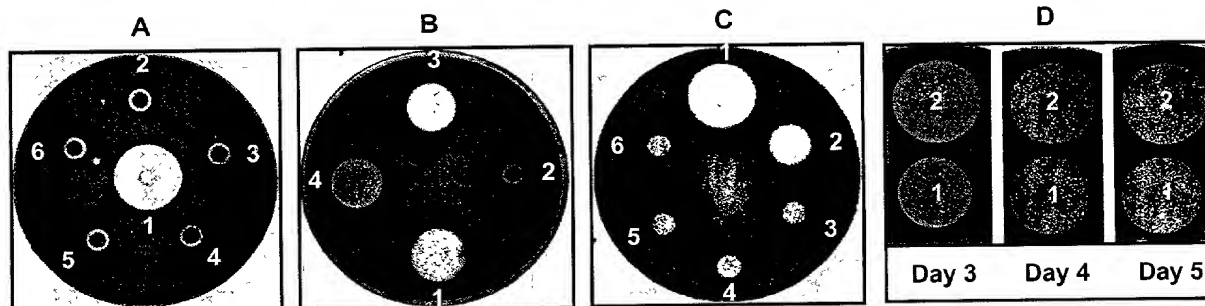
Fig. 9A   Fig. 9B   Fig. 9C   Fig. 9D

PATENT APPLICATION DECLARATION

(Attorney's Docket No.: 2500.125US2)

Each of the Applicants named below hereby declares as follows:

1. My residence, post office address and country of citizenship given below are true and correct.

2. I believe I am the original, first and joint inventor of the subject matter which is claimed and for which a patent is sought in the patent application entitled "GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR ANTIBIOTIC C-1027," Serial No. _____, filed January 5, 2000, and I have reviewed and understand the contents of the specification, including its claims.

3. I acknowledge my duty to disclose to the Office all information known to me to be material to patentability of this application, in accordance with 37 C.F.R. Section 1.56, which is defined on the attached page.

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Date: _____

_____

                                      Ben Shen

Residence and   1842 Rushmore Lane
Post Office Address: Davis, California 95616
                                        (Citizenship: People's Republic of China)

Date: _____

_____

                                        Wen Liu

Residence and   Institute of Medicinal Biotechnology
Post Office Address: Tiantan, Beijing, 100005, China
                                        (Citizenship: Peoples Republic of China)

Date: _____   _____
                           Steven D. Christenson
         Residence and   1079 Monarch Lane
Post Office Address: Davis, California 95616
                           (Citizenship: United States)

Date: _____   _____
                           Scott Standage
         Residence and   63 Tudor Road
Post Office Address: Bornet, Herts, EN5 5NW, United Kingdom
                           (Citizenship: United Kingdom)

Section 1.56 Duty to Disclose Information Material to Patentability.

(a) A patent by its very nature is affected with a public interest. The public interest is best served, and the most effective patent examination occurs when, at the time an application is being examined, the Office is aware of and evaluates the teachings of all information material to patentability. Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability as defined in this section. The duty to disclose information exists with respect to each pending claim until the claim is cancelled or withdrawn from consideration, or the application becomes abandoned. Information material to the patentability of a claim that is cancelled or withdrawn from consideration need not be submitted if the information is not material to the patentability of any claim remaining under consideration in the application. There is no duty to submit information which is not material to the patentability of any existing claim. The duty to disclose all information known to be material to patentability is deemed to be satisfied if all information known to be material to patentability of any claim issued in a patent was cited by the Office or submitted to the Office in the manner prescribed by §§ 1.97(b)-(d) and 1.98. However, no patent will be granted on an application in connection with which fraud on the Office was practiced or attempted or the duty of disclosure was violated through bad faith or intentional misconduct. The Office encourages applicants to carefully examine:

(1) prior art cited in search reports of a foreign patent office in a counterpart application, and

(2) the closest information over which individuals associated with the filing or prosecution of a patent application believe any pending claim patentably defines, to make sure that any material information contained therein is disclosed to the Office.

(b) Under this section, information is material to patentability when it is not cumulative to information already of record or being made of record in the application, and

(1) It establishes, by itself or in combination with other information, a prima facie case of unpatentability of a claim; or

(2) It refutes, or is inconsistent with, a position the applicant takes in:

(i) Opposing an argument of unpatentability relied on by the Office, or

(ii) Asserting an argument of patentability.

A prima facie case of unpatentability is established when the information compels a conclusion that a claim is unpatentable under the preponderance of evidence, burden-of-proof standard, giving each term in the claim its broadest reasonable construction consistent with the specification, and before any consideration is given to evidence which may be submitted in an attempt to establish a contrary conclusion of patentability.

(c) Individuals associated with the filing or prosecution of a patent application within the meaning of this section are:

(1) Each inventor named in the application;

(2) Each attorney or agent who prepares or prosecutes the application; and

(3) Every other person who is substantively involved in the preparation or prosecution of the application and who is associated with the inventor, with the assignee or with anyone to whom there is an obligation to assign the application.

(d) Individuals other than the attorney, agent or inventor may comply with this section by disclosing information to the attorney, agent, or inventor.

GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR ANTIBIOTICC-1027

SEQUENCE LISTING

SEQ ID No. 1. C-1027 gene cluster DNA sequence from 1 to 42,000, ORF-(-7) to ORF-26

```
          GTCGACTCTAGAGGATCCCGGGTGCGGAGTAGGGGTTACGGACGAAGGAGGGGTGCCCGG
        1 ---------+---------+---------+---------+---------+---------+ 60
          CAGCTGAGATCTCCTAGGGCCCACGCCTCATCCCCAATGCCTGCTTCCTCCCCACGGGCC
   -7-*             *  L  I  G  P  A  S  Y  P  N  R  V  F  S  P  H  G  -

          CGACGCCTGCGGCGAAGGGCGGTTCCTTGAGTTCGAGGCCGGTGGCGAGGACGACGTGGT
       61 ---------+---------+---------+---------+---------+---------+ 120
          GCTGCGGACGCCGCTTCCCGCCAAGGAACTCAAGCTCCGGCCACCGCTCCTGCTGCACCA
   -7     A  V  G  A  A  F  P  P  E  K  L  E  L  G  T  A  L  V  V  H  -

          CCGCGTCGAGGATCTGCGTGTCGGGGAGCGGCCCAGGGCGCAGCCCCTCGGTCAGGTACG
      121 ---------+---------+---------+---------+---------+---------+ 180
          GGCGCAGCTCCTAGACGCACAGCCCCTCGCCGGGTCCCGCGTCGGGGAGCCAGTCCATGC
   -7     D  A  D  L  I  Q  T  D  P  L  P  G  P  R  L  G  E  T  L  Y  -

          GGGTGAGGCCCCTGACGGTCACCTCGAAGCAGCGGTCGTGGGACCGGGCGTCGAGCGCCT
      181 ---------+---------+---------+---------+---------+---------+ 240
          CCCACTCCGGGGACTGCCAGTGGAGCTTCGTCGCCAGCACCCTGGCCCGCAGCTCGCGGA
   -7     P  T  L  G  R  V  T  V  E  F  C  R  D  H  S  R  A  D  L  A  -

          CCCCGTCCGCTTCCACAAGGACGACGCCGGGACAGGACTCCCGTGCGGCCTCGACCAGTC
      241 ---------+---------+---------+---------+---------+---------+ 300
          GGGGCAGGCGAAGGTGTTCCTGCTGCGGCCCTGTCCTGAGGGCACGCCGGAGCTGGTCAG
   -7     E  G  D  A  E  V  L  V  V  G  P  C  S  E  R  A  A  E  V  L  -

          GGGCGTCGAGGTAGTCCTGGAAGATGCGGCGGGGGGCGGGGCCCTGTTCGGTGAACTTCC
      301 ---------+---------+---------+---------+---------+---------+ 360
          CCCGCAGCTCCATCAGGACCTTCTACGCCGCCCCCCGCCCCGGGACAAGCCACTTGAAGG
   -7     R  A  D  L  Y  D  Q  F  I  R  R  P  A  P  G  Q  E  T  F  K  -

          ACGAAGCCCAGCGCCGGGGGCCAGTCGCGCCGGTCGGCCTCCTGGTTGGCCCAGTTGATGA
      361 ---------+---------+---------+---------+---------+---------+ 420
          TGCTTCGGGTCGCGGCCCCGGTCAGCGCGGCCAGCCGGAGGACCAACCGGGTCAACTACT
   -7     W  S  A  W  R  R  P  W  D  R  R  D  A  E  Q  N  A  W  N  I  -

          AGTCGAGCACGTCCTCGCGGAACACCGACATCCTGCCGGCCTGGATATTGAAGACGTGGT
      421 ---------+---------+---------+---------+---------+---------+ 480
          TCAGCTCGTGCAGGAGCGCCTTGTGGCTGTAGGACGGCCGGACCTATAACTTCTGCACCA
   -7     F  D  L  V  D  E  R  F  V  S  M  R  G  A  Q  I  N  F  V  H  -

          CCCAGGGGTTGCCGTCACGGTGATAGGCGACGCCGGCCGAGCGGTAgGCGGCGCGCCGCT
      481 ---------+---------+---------+---------+---------+---------+ 540
          GGGTCCCCAACGGCAGTGCCACTATCCGCTGCGGCCGGCTCGCCATcCGCCGCGCGGCGA
   -7     D  W  P  N  G  D  R  H  Y  A  V  G  A  S  R  Y  A  A  R  -

          CCAGGAGGACGACTTCCAGCGGTCTTCTCGCGAAATGAAGCAGGCGTATCGCGGTCGCCG
      541 ---------+---------+---------+---------+---------+---------+ 600
          GGTCCTCCTGCTGAAGGTCGCCAGAAGAGCGCTTTACTTCGTCCGCATAGCGCCAGCGGC
   -7     E  L  L  V  V  E  L  P  R  R  A  F  H  L  L  R  I  A  T  A  -

          TGCCTGCCAGGCCCGCCCCTACGACCAGCACCCTGGGGCGCGCACCCGTCATGCCCATGA
      601 ---------+---------+---------+---------+---------+---------+ 660
          ACGGACGGTCCGGGCGGGGATGCTGGTCGTGGGACCCCGCGCGTGGGCAGTACGGGTACT
   -7-<    T  G  A  L  G  A  G  V  V  L  V  R  P  R  A  G  T  M  G  M  -
```

```
        AGCCTCCCCCGCTGACTCAGGGCGgCGCGTCGCGCGCTCCCGTCGGTGTCCTCGCTGACT
  661   ---------+---------+---------+---------+---------+---------+  720
        TCGGAGGGGGCGACTGAGTCCCGCcGCGCAGCGCGCGAGGGCAGCCACAGGAGCGACTGA

        GGAAGTTCCCTGACCTGGCGTCAACTCCACTGATCCGTAAGGGGATCGCGGGAGTGGATA
  721   ---------+---------+---------+---------+---------+---------+  780
        CCTTCAAGGGACTGGACCGCAGTTGAGGTGACTAGGCATTCCCCTAGCGCCCTCACCTAT

        CGGGTCAGGTCGTGCACGATCGTGGCACCAGACAGATCACCACGTCGATAGGCACTCGTG
  781   ---------+---------+---------+---------+---------+---------+  840
        GCCCAGTCCAGCACGTGCTAGCACCGTGGTCTGTCTAGTGGTGCAGCTATCCGTGAGCAC

        AGCCGCGCCCGGGGCTCGACGGGGCGGGGCACCGGCAGGGGCGGCCGCGTGATCAGCCGG
  841   ---------+---------+---------+---------+---------+---------+  900
        TCGGCGCGGGCCCCGAGCTGCCCCGCCCCGTGGCCGTCCCCGCCGGCGCACTAGTCGGCC

        AGCCTGTCCGGGGGCGTGCGTGCGGGGCGTCAGCTGTCGATGTCGGGAACGCCAGGGACG
  901   ---------+---------+---------+---------+---------+---------+  960
        TCGGACAGGCCCCCGCACGCACGCCCCGCAGTCGACAGCTACAGCCCTTGCGGTCCCTGC
  -6-*                                          *  S  D  I  D  P  V  G  P  V  -

        TCGATCTCGGTGCGGGCGTAGTGGTTGAAGTAGTTGGTGTAGAGGTTCACGGCCACGTGG
  961   ---------+---------+---------+---------+---------+---------+  1020
        AGCTAGAGCCACGCCCGCATCACCAACTTCATCAACCACATCTCCAAGTGCCGGTGCACC
  -6     D  I  E  T  R  A  Y  H  N  F  Y  N  T  Y  L  N  V  A  V  H  -

        ACGAAGACCTCGGCGAGCTCGGTGTCCGTCCATCCCTGTGCCACGGCCGCGTTCCACGAG
 1021   ---------+---------+---------+---------+---------+---------+  1080
        TGCTTCTGGAGCCGCTCGAGCCACAGGCAGGTAGGGACACGGTGCCGGCGCAAGGTGCTC
  -6     V  F  V  E  A  L  E  T  D  T  W  G  Q  A  V  A  A  N  W  S  -

        GCGTCAGACGCCTCGCCCACTTCGCCGGCGATCTCCCTGGCCACCTGGACCAGTGCTTCG
 1081   ---------+---------+---------+---------+---------+---------+  1140
        CGCAGTCTGCGGAGCGGGTGAAGCGGCCGCTAGAGGGACCGGTGGACCTGGTCACGAAGC
  -6     A  D  S  A  E  G  V  E  G  A  I  E  R  A  V  Q  V  L  A  E  -

        AGCTTCACGTCGTCGCCGGGCGTCCCCCGGCGAATCGCCACGGTCTCCTCCAGCGTGAAA
 1141   ---------+---------+---------+---------+---------+---------+  1200
        TCGAAGTGCAGCAGCGGCCCGCAGGGGGCCGCTTAGCGGTGCCAGAGGAGGTCGCACTTT
  -6     L  K  V  D  D  G  P  T  G  R  R  I  A  V  T  E  E  L  T  F  -

        CCCGCGACCTTCGCCGACACCGTGTGCGCCGCCTGGCAGTACGCGCACGCGTCGACCGCG
 1201   ---------+---------+---------+---------+---------+---------+  1260
        GGGCGCTGGAAGCGGCTGTGGCACACGCGGCGGACCGTCATGCGCGTGCGCAGCTGGCGC
  -6     G  A  V  K  A  S  V  T  H  A  A  Q  C  Y  A  C  A  D  V  A  -

        CCCACGGCGAGGGCGATCGCCTCGCGTGTGCGGGCGTCGAACGTTCCATGTTCGGCGACG
 1261   ---------+---------+---------+---------+---------+---------+  1320
        GGGTGCCGCTCCCGCTAGCGGAGCGCACACGCCCGCAGCTTGCAAGGTACAAGCCGCTGC
  -6     G  V  A  L  A  I  A  E  R  T  R  A  D  F  T  G  H  E  A  V  -

        GCTCCGGTGATCGCGGCGTAGGTTTCCAGGACCACGGGGGAATGGGCCATTCCCCCGTGG
 1321   ---------+---------+---------+---------+---------+---------+  1380
        CGAGGCCACTAGCGCCGCATCCAAAGGTCCTGGTGCCCCCTTACCCGGTAAGGGGGCACC
  -6     A  G  T  I  A  A  Y  T  E  L  V  V  P  S  H  A  M  G  G  H  -

        ATGTTGAGCACTCGCCCGAACCGCTTCTCCAGTCGGCGCAGGATGTCTCCGCCGGCTGCG
 1381   ---------+---------+---------+---------+---------+---------+  1440
        TACAACTCGTGAGCGGGCTTGGCGAAGAGGTCAGCCGCGTCCTACAGAGGCGGCCGACGC
  -6     I  N  L  V  R  G  F  R  K  E  L  R  R  L  I  D  G  A  A  -

        GGTGCGGTGTCGATGGTGTGGACGGGAATCCGCGGCATGGGAATGCCTCTCCTCGTAGTG
 1441   ---------+---------+---------+---------+---------+---------+  1500
        CCACGCCACAGCTACCACACCAGCCCTTAGGCGCCGTACCCTTACGGAGAGGAGCATCAC
  -6-<   P  A  T  D  I  T  H  V  P  I  R  P  M
```

2

```
         ATGGGAGTTCCTCGTCCCTCCAGTCTGCCCAAGCACCTCCCCCGGTGAGCTGTCCCGGCC
1501     ---------+---------+---------+---------+---------+---------+ 1560
         TACCCTCAAGGAGCAGGGAGGTCAGACGGGTTCGTGGAGGGGGCCACTCGACAGGGCCGG


         GCCCTCCGGCCCCTTCTAGGCAGGTCGCCCGGTGGTGCGGCCCCAGGACGTCACCTCGCC
1561     ---------+---------+---------+---------+---------+---------+ 1620
         CGGGAGGCCGGGGAAGATCCGTCCAGCGGGCCACCACGCCGGGGTCCTGCAGTGGAGCGG


         GCACCACCGGGAGCCCCGAGGGGCGAGGTCAGAGGCCGAGCACCTCCTCGGCCAGGGCGG
1621     ---------+---------+---------+---------+---------+---------+ 1680
         CGTGGTGGCCCTCGGGGCTCCCCGCTCCAGTCTCCGGCTCGTGGAGGAGCCGGTCCCGCC
-5-*                                            *  L  G  L  V  E  E  A  L  A  -


         TGCCCCGAACACGGGCCTCGATCTTGGCGAAGGCCAGGTCGCGTGTGGTGGAGGTGTCGT
1681     ---------+---------+---------+---------+---------+---------+ 1740
         ACGGGGCTTGTGCCCGGAGCTAGAACCGCTTCCGGTCCAGCGCACACCACCTCCACAGCA
-5        T  G  R  V  R  A  E  I  K  A  F  A  L  D  R  T  T  S  T  D  -


         CGGCGAACGGGGAGAAGCCGCAGTCGTCGCAGGTTCCCAGTTGCTCGACGGGGATGTAGC
1741     ---------+---------+---------+---------+---------+---------+ 1800
         GCCGCTTGCCCCTCTTCGGCGTCAGCAGCGTCCAAGGGTCAACGAGCTGCCCCTACATCG
-5        D  A  F  P  S  F  G  C  D  D  C  T  G  L  Q  E  V  P  I  Y  -


         GGGCGGCGAGCAGGATGCGGTCGCGTACCTGCTCGGGGGTCTCGACCACTGGGTCGATCG
1801     ---------+---------+---------+---------+---------+---------+ 1860
         CCCGCCGCTCGTCCTACGCCAGCGCATGGACGAGCCCCCAGAGCTGGTGACCCAGCTAGC
-5        R  A  A  L  L  I  R  D  R  V  Q  E  P  T  E  V  V  P  D  I  -


         GGTCGGTCACCCCGAGGAAGACGCGGGCGGCAGGGGGCAGGTGGTCACGGACGATGCTCA
1861     ---------+---------+---------+---------+---------+---------+ 1920
         CCAGCCAGTGGGGCTCCTTCTGCGCCCGCCGTCCCCCGTCCACCAGTGCCTGCTACGAGT
-5        P  D  T  V  G  L  F  V  R  A  A  P  P  L  H  D  R  V  I  S  -


         GGACCCGCTCGGGGTCCGCTTCGCCGGCCAGTTCGAGATAGAAGTTGCCCGCCTTGAGCT
1921     ---------+---------+---------+---------+---------+---------+ 1980
         CCTGGGCGAGCCCCAGGCGAAGCGGCCGGTCAAGCTCTATCTTCAACGGGCGGAACTCGA
-5        L  V  R  E  P  D  A  E  G  A  L  E  L  Y  F  N  G  A  K  L  -


         GGAAGAGCTTGGGCAGCAGTTCGGCGTAGTCGATGTCGAGGCTGTGCGTGGAGTCCTGGT
1981     ---------+---------+---------+---------+---------+---------+ 2040
         CCTTCTCGAACCCGTCGTCAAGCCGCATCAGCTACAGCTCCGACACGCACCTCAGGACCA
-5        Q  F  L  K  P  L  L  E  A  Y  D  I  D  L  S  H  T  S  D  Q  -


         CGCCGCCGGGGCAGGTGTGTACGCCGATGCGGGCGGTTTCCTCGGCGCTGAAGCGCCCCA
2041     ---------+---------+---------+---------+---------+---------+ 2100
         GCGGCGGCCCCGTCCACACATGCGGCTACGCCCGCCAAAGGAGCCGCGACTTCGCGGGGT
-5        D  G  G  P  C  T  H  V  G  I  R  A  T  E  E  A  S  F  R  G  -


         GGACTTCGTTGTTGAGGGCGATGAAGTCGTCGAGGACGCCGCCGCTGGGGTCGAGCTTGA
2101     ---------+---------+---------+---------+---------+---------+ 2160
         CCTGAAGCAACAACTCCCGCTACTTCAGCAGCTCCTGCGGCGGCGACCCCAGCTCGAACT
-5        L  V  E  N  N  L  A  I  F  D  D  L  V  G  G  S  P  D  L  K  -


         GGGACAGCCGCCCCTCGGTGAAGTCGAGCTGGACCACGTGTGCCCCCGCGTCCAGGCAGC
2161     ---------+---------+---------+---------+---------+---------+ 2220
         CCCTGTCGGCGGGGAGCCACTTCAGCTCGACCTGGTGCACACGGGGGCGCAGGTCCGTCG
-5        L  S  L  R  G  E  T  F  D  L  Q  V  V  H  A  G  A  D  L  C  -


         CTCGGATGTCGGCTTCGGCCTCGTCGGCGAGGTCGCGCAGGAACTGCTCGCGGGGGTAGC
2221     ---------+---------+---------+---------+---------+---------+ 2280
         GAGCCTACAGCCGAAGCCGGAGCAGCCGCTCCAGCGCGTCCTTGACGAGCGCCCCCATCG
-5        G  R  I  D  A  E  A  E  D  A  L  D  R  L  F  Q  E  R  P  Y  -


         CCTCGATGGGAGTGGCGGGGTAGAGGAGGCTGAGGGCGGAGGGTGCGATGACCGCCTGCT
2281     ---------+---------+---------+---------+---------+---------+ 2340
```

3

```
       GGAGCTACCCTCACCGCCCCATCTCCTCCGACTCCCGCCTCCCACGCTACTGGCGGACGA
  -5     G  E  I  P  T  A  P  Y  L  L  S  L  A  S  P  A  I  V  A  Q   -

       TCAGGGGGCGGTCCGTGAGCTGCCGTGCGGCGCGCAGATAGGTTTCGGCCCGCACCTGGT
 2341  ---------+---------+---------+---------+---------+---------+ 2400
       AGTCCCCCGCCAGGCACTCGACGGCACGCCGCGCGTCTATCCAAAGCCGGGCGTGGACCA
  -5     K  L  P  R  D  T  L  Q  R  A  A  R  L  Y  T  E  A  R  V  Q   -

       AGCGGAAGGGCCCTTGGGTGATGCTGGGGAGCTGCCGGGTGTGCCCGTCTGCGAAGGGGA
 2401  ---------+---------+---------+---------+---------+---------+ 2460
       TCGCCTTCCCGGGAACCCACTACGACCCCTCGACGGCCCACACGGGCAGACGCTTCCCCT
  -5     Y  R  F  P  G  Q  T  I  S  P  L  Q  R  T  H  G  D  A  F  P   -

       TGACAGCGCCGTCGGGCGAGAGGGTGTCGAGGCCGGTCACGGGGTAGGTGGCGAAGCTCG
 2461  ---------+---------+---------+---------+---------+---------+ 2520
       ACTGTCGCGGCAGCCCGCTCTCCCACAGCTCCGGCCAGTGCCCCATCCACCGCTTCGAGC
  -5     I  V  A  G  D  P  S  L  T  D  L  G  T  V  P  Y  T  A  F  S   -

       GCTTGGACTGTTCACCGTCCACGAGGACGGGGCTGCCGACTCGTTCCAGTCGTGTCAGGG
 2521  ---------+---------+---------+---------+---------+---------+ 2580
       CGAACCTGACAAGTGGCAGGTGCTCCTGCCCCGACGGCTGAGCAAGGTCAGCACAGTCCC
  -5     P  K  S  Q  E  G  D  V  L  V  P  S  G  V  R  E  L  R  T  L   -

       TGTCCGCGACGGCCTGTTCCTGCTGTTTGGCCAGGTCCGTGGCGTCCAGGGTTCCCTGGG
 2581  ---------+---------+---------+---------+---------+---------+ 2640
       ACAGGCGCTGCCGGACAAGGACGACAAACCGGTCCAGGCACCGCAGGTCCCAAGGGACCC
  -5     T  D  A  V  A  Q  E  Q  Q  K  A  L  D  T  A  D  L  T  G  Q   -

       CATGCGCGGCAAGGGCGTGCAGGAGTGTCGCGGAGCGCGGAAGGCTGCCGATCGGCTCAG
 2641  ---------+---------+---------+---------+---------+---------+ 2700
       GTACGCGCCGTTCCCGCACGTCCTCACAGCGCCTCGCGCCTTCCGACGGCTAGCCGAGTC
  -5     A  H  A  A  L  A  H  L  L  T  A  S  R  P  L  S  G  I  P  E   -

       TGGCGATGGTCATGGCCGAAGAGTAGGGAAGAGGCTGGGTTTCGAACCACCGCAAAGCTT
 2701  ---------+---------+---------+---------+---------+---------+ 2760
       ACCGCTACCAGTACCGGCTTCTCATCCCTTCTCCGACCCAAAGCTTGGTGGCGTTTCGAA
  -5-<   T  A  I  T  M  -

       TGATTGCCGCTTTTTTCAGGGGAAGTTGATGCGAAGTCGCCGAGCGGCGGAACGTGCTGAT
 2761  ---------+---------+---------+---------+---------+---------+ 2820
       ACTAACGGCGAAAAAGTCCCCTTCAACTACGCTTCAGCGGCTCGCCGCCTTGCACGACTA

       GTATGGGGGGCGGGAGGAGCCTGCGGGGTTCTAGGAGCCGGTCGCGGCCACGGTGGAGGA
 2821  ---------+---------+---------+---------+---------+---------+ 2880
       CATACCCCCCGCCCTCCTCGGACGCCCCAAGATCCTCGGCCAGCGCCGGTGCCACCTCCT
  -4-*                                       *  S  G  T  A  A  V  T  S  S   -

       GGTGCCCAGCTGGGAGCGGGGGGTCTTTTCGCCGACGCGGTTGGGCTCGATGGTGCGGGG
 2881  ---------+---------+---------+---------+---------+---------+ 2940
       CCACGGGTCGACCCTCGCCCCCCAGAAAAGCGGCTGCGCCAACCCGAGCTACCACGCCCC
  -4     T  G  L  Q  S  R  P  T  K  E  G  V  R  N  P  E  I  T  R  P  -

       GTCGACGGCCTCTCCGGGGGCACCTTGCCGGTAGACGCCTTCGGGGTCGGAGTCCCGGTC
 2941  ---------+---------+---------+---------+---------+---------+ 3000
       CAGCTGCCGGAGAGGCCCCCGTGGAACGGCCATCTGCGGAAGCCCCAGCCTCAGGGCCAG
  -4     D  V  A  E  G  P  A  G  Q  R  Y  V  G  E  P  D  S  D  R  D  -

       ATGGGGGAGCAGGAAGAAGACCCGGCGCCGGTACAGACCGCTGTCCGGGTCCGCTTCGGC
 3001  ---------+---------+---------+---------+---------+---------+ 3060
       TACCCCCTCGTCCTTCTTCTGGGCCGCGGCCATGTCTGGCGACAGGCCCAGGCGAAGCCG
  -4     H  P  L  L  F  F  V  R  R  R  Y  L  G  S  D  P  D  A  E  A  -

       GTCGGCCCCGAGTTCGATGTAGCCGATCATGCGGCCGTCGCGGGCGTAGCGCGGCTTGTT
 3061  ---------+---------+---------+---------+---------+---------+ 3120
       CAGCCGGGGGCTCAAGCTACATCGGCTAGTACGCCGGCAGCGCCCGCATCGCGCCGAACAA
  -4     D  A  G  L  E  I  Y  G  I  M  R  G  D  R  A  Y  R  P  K  N  -
```

4

```
                    CTTGCGCCGGGGGGTCTTGTCCAGGGCCTGGCGGACGTAGTCGAGTCCCTCGGGATCTTC
          3121     ---------+---------+---------+---------+---------+---------+ 3180
                    GAACGCGGCCCCCCAGAACAGGTCCCGGACCGCCTGCATCAGCTCAGGGAGCCCTAGAAG
          -4           K  R  R  P  T  K  D  L  A  Q  R  V  Y  D  L  G  E  P  D  E -

                    GAGCCACACGACCTTCGCCTCGTGAACGAGATCGCTGTCGGTCAGTAGCGAGCTCATGGC
          3181     ---------+---------+---------+---------+---------+---------+ 3240
                    CTCGGTGTGCTGGAAGCGGAGCACTTGCTCTAGCGACAGCCAGTCATCGCTCGAGTACCG
          -4-<         L  W  V  V  K  A  E  H  V  L  D  S  D  T  L  L  S  S  M -

                    GGCGACCTCTCCTTCGTCGGCGTGCACCGGGTGGGGAAGCGGTGCCTGCGTGATGTGTGT
          3241     ---------+---------+---------+---------+---------+---------+ 3300
                    CCGCTGGAGAGGAAGCAGCCGCACGTGGCCCACCCCTTCGCCACGGACGCACTACACACA

                    TCGTCTGCGGCGGTGGGCCGCAGTGGTGCGGACCGCCCGTGGTGCCGGTTCTCGGCCAAA
          3301     ---------+---------+---------+---------+---------+---------+ 3360
                    AGCAGACGCCGCCACCCGGCGTCACCACGCCTGGCGGGCACCACGGCCAAGAGCCGGTTT

                    GCACGGGCAGGTACGTCCTGGGGCACTCACATCGTAGATGGGGTCCGCTTCCGCAGGGCA
          3361     ---------+---------+---------+---------+---------+---------+ 3420
                    CGTGCCCGTCCATGCAGGACCCCGTGAGTGTAGCATCTACCCCAGGCGAAGGCGTCCCGT

                    GTGCCTCCGGTCGGAGGACGTTCATTCGTCGGCTGCCAGAGCGAGGTTGGGGTAGAACTT
          3421     ---------+---------+---------+---------+---------+---------+ 3480
                    CACGGAGGCCAGCCTCCTGCAAGTAAGCAGCCGACGGTCTCGCTCCAACCCCATCTTGAA
          -3-*                                   *  E  D  A  A  L  A  L  N  P  Y  F  K -

                    CCGGCCGTTGGATTTGATCATGTCGGCAGGTGAGGCGAGGCCCACTTCCTGGCGGACCCG
          3481     ---------+---------+---------+---------+---------+---------+ 3540
                    GGCCGGCAACCTAAACTAGTACAGCCGTCCACTCCGCTCCGGGTGAAGGACCGCCTGGGC
          -3           R  G  N  S  K  I  M  D  A  P  S  A  L  G  V  E  Q  R  V  R -

                    GGTGGCGAAGGCACGGGCGGTCCCGGGGCGGATGCCTTCACTGTGTGCGCACCAGGTGCT
          3541     ---------+---------+---------+---------+---------+---------+ 3600
                    CCACCGCTTCCGTGCCCGCCAGGGCCCCGCCTACGGAAGTGACACACGCGTGGTCCACGA
          -3           T  A  F  A  R  A  T  G  P  R  I  G  E  S  H  A  C  W  T  S -

                    GTAGGACGTGTAGAGAAGGCCCTGTTCGACGCGTAGCTCGCTGTTCTCGGGGTCGTGGAG
          3601     ---------+---------+---------+---------+---------+---------+ 3660
                    CATCCTGCACATCTCTTCCGGGACAAGCTGCGCATCGAGCGACAAGAGCCCCAGCACCTC
          -3           Y  S  T  Y  L  L  G  Q  E  V  R  L  E  S  N  E  P  D  H  L -

                    GCAGCACTCGGCGAGGAAGCGGCCGATGTGGTCCTCGGTGTTCGCGTATGCGCTGGTGGC
          3661     ---------+---------+---------+---------+---------+---------+ 3720
                    CGTCGTGAGCCGCTCCTTCGCCGGCTACACCAGGAGCCACAAGCGCATACGCGACCACCG
          -3           C  C  E  A  L  F  R  G  I  H  D  E  T  N  A  Y  A  S  T  A -

                    GATGCGGACCCGGTCGGGGCCGGCGAGTGTGTCGCGGGTGGCGAGGTAGCGGCGGGCCCC
          3721     ---------+---------+---------+---------+---------+---------+ 3780
                    CTACGCCTGGGCCAGCCCCGGCCGCTCACACAGCGCCCACCGCTCCATCGCCGCCCGGGG
          -3           I  R  V  R  D  P  G  A  L  T  D  R  T  A  L  Y  R  R  A  G -

                    TTCGGTGAGCCAGTGCAGGATCCCGGGGCCCTCGTCCTGGACGAGTTCGACAGCCAGGTT
          3781     ---------+---------+---------+---------+---------+---------+ 3840
                    AAGCCACTCGGTCACGTCCTAGGGCCCCGGGAGCAGGACCTGCTCAAGCTGTCGGTCCAA
          -3           E  T  L  W  H  L  I  G  P  G  E  D  Q  V  L  E  V  A  L  N -

                    GTCGATCTTGCGTTCGTCGGGGACGATCCGTTCGAAGGGCAGGAGGCGGATGCGGCGCCA
          3841     ---------+---------+---------+---------+---------+---------+ 3900
                    CAGCTAGAACGCAAGCAGCCCCTGCTAGGCAAGCTTCCCGTCCTCCGCCTACGCCGCGGT
          -3           D  I  K  R  E  D  P  V  I  R  E  F  P  L  L  R  I  R  R  W -

                    GAAGGCGAAGCCGCCGGTGGAGACCTCGGGGCGGTGGTTGCCCAGCAGCCACAGCTTGTG
          3901     ---------+---------+---------+---------+---------+---------+ 3960
                    CTTCCGCTTCGGCGGCCACCTCTGGAGCCCCGCCACCAACGGGTCGTCGGTGTCGAACAC
```

5

```
-3          F  A  F  G  G  T  S  V  E  P  R  H  N  G  L  L  W  L  K  H -

            CGTGGGTGTGAAGGAGAAATAGTCCTGCCGCATGCGGCGGGCCTTGATCTTGTCACCGCC
      3961  ---------+---------+---------+---------+---------+---------+  4020
            GCACCCACACTTCCTCTTTATCAGGACGGCGTACGCCGCCCGGAACTAGAACAGTGGCGG
-3          T  P  T  F  S  F  Y  D  Q  R  M  R  R  A  K  I  K  D  G  G -

            GGTCAGCAGGCGGACGCGCGCCTCGTCGAAGCGGTCGTTGGGCTTGAGCTCGCTGCACAC
      4021  ---------+---------+---------+---------+---------+---------+  4080
            CCAGTCGTCCGCCTGCGCGCGGAGCAGCTTCGCCAGCAACCCGAACTCGAGCGACGTGTG
-3             T  L  L  R  V  R  A  E  D  F  R  D  N  P  K  L  E  S  C  V -

            GATGAGGCGGCGGCCGTGGAGTTCGGTGAGCTCGGTGGAGTGTTCGGAGTATGCGCCACG
      4081  ---------+---------+---------+---------+---------+---------+  4140
            CTACTCCGCCGCCGGCACCTCAAGCCACTCGAGCCACCTCACAAGCCTCATACGCGGTGC
-3             I  L  R  R  G  H  L  E  T  L  E  T  S  H  E  S  Y  A  G  R -

            GTCCATGAGGAAACCCGGCGGGGCTGCGTCGGCGTAGTCGCCGAGAATCTGGATCATCAC
      4141  ---------+---------+---------+---------+---------+---------+  4200
            CAGGTACTCCTTTGGGCCGCCCCGACGCAGCCGCATCAGCGGCTCTTAGACCTAGTAGTG
-3          D  M  L  F  G  P  P  A  A  D  A  Y  D  G  L  I  Q  I  M  V -

            GTCGAGGAGAACGGATTTGCCGTTCTTTCCCTGGCCGTGGAGAAAGGGCAGCACCTGCGC
      4201  ---------+---------+---------+---------+---------+---------+  4260
            CAGCTCCTCTTGCCTAAACGGCAAGAAAGGGACCGGCACCTCTTTCCCGTCGTGGACGCG
-3          D  L  L  V  S  K  G  N  K  G  Q  G  H  L  F  P  L  V  Q  A -

            CCCGACGTCACCGGTGATGGAGTAGCCGAGAAGGAGGTGGAGGAAGTCGATCATCTCCCG
      4261  ---------+---------+---------+---------+---------+---------+  4320
            GGGCTGCAGTGGCCACTACCTCATCGGCTCTTCCTCCACCTCCTTCAGCTAGTAGAGGGC
-3             G  V  D  G  T  I  S  Y  G  L  L  L  H  L  F  D  I  M  E  R -

            CCCTTCGGCGTCACTGCCGAAGGTGTCTTCGAGGAAACGGTGCCAGCGGGGGGTGGGGAT
      4321  ---------+---------+---------+---------+---------+---------+  4380
            GGGAAGCCGCAGTGACGGCTTCCACAGAAGCTCCTTTGCCACGGTCGCCCCCCACCCCTA
-3             G  E  A  D  S  G  F  T  D  E  L  F  R  H  W  R  P  T  P  I -

            GTCCTGGGGGGAGGCGCTGGTGGCGCGGGAGTGGAAGTCCCGGGTGGGGTCGGGCTTGCG
      4381  ---------+---------+---------+---------+---------+---------+  4440
            CAGGACCCCCCTCCGCGACCACCGCGCCCTCACCTTCAGGGCCCACCCCAGCCCGAACGC
-3             D  Q  P  S  A  S  T  A  R  S  H  F  D  R  T  P  D  P  K  R -

            CATACGGCCGTTGCGGAGGTCGACCACTCCGTCAGGGGTGCACAGGGCGTAGGGGTCTCC
      4441  ---------+---------+---------+---------+---------+---------+  4500
            GTATGCCGGCAACGCCTCCAGCTGGTGAGGCAGTCCCCACGTGTCCCGCATCCCCAGAGG
-3             M  R  G  N  R  L  D  V  V  G  D  P  T  C  L  A  Y  P  D  G -

            GTCGAGGGTGTCGGGATCGAGGGAGAGGTCGGGAGAGGCCTTTGCCTGGGTGAGGAGCGC
      4501  ---------+---------+---------+---------+---------+---------+  4560
            CAGCTCCCACAGCCCTAGCTCCCTCTCCAGCCCTCTCCGGAAACGGACCCACTCCTCGCG
-3          D  L  T  D  P  D  L  S  L  D  P  S  A  K  A  Q  T  L  L  A -

            CTTCATACCGGTCGTCGACAGGGTGCGGCGTTTGTGGTGGTGCAGTTCCCGGTCGGTGAA
      4561  ---------+---------+---------+---------+---------+---------+  4620
            GAAGTATGGCCAGCAGCTGTCCCACGCCGCAAACACCACCACGTCAAGGGCCAGCCACTT
-3          K  M  G  T  T  S  L  T  R  R  K  H  H  H  L  E  R  D  T  F -

            CAGCCCGCGGGGATCGCTGCCGGGCATCTCCTCCGCCATCTCTCCGGCAGCCCACAGGGC
      4621  ---------+---------+---------+---------+---------+---------+  4680
            GTCGGGCGCCCCTAGCGACGGCCCGTAGAGGAGGCGGTAGAGAGGCCGTCGGGTGTCCCG
-3          L  G  R  P  D  S  G  P  M  E  E  A  M  E  G  A  A  W  L  A -

            AGCTTTCTCGCCTCCGGCCCGCTTCCACCGGTAGCCGTCCCAGGAGTACCAGCCCAGGCC
      4681  ---------+---------+---------+---------+---------+---------+  4740
            TCGAAAGAGCGGAGGCCGGGCGAAGGTGGCCATCGGCAGGGTCCTCATGGTCGGGTCCGG
```

6

```
-3          A  K  E  G  G  A  R  K  W  R  Y  G  D  W  S  Y  W  G  L  G -

      CTCCACGTGCCGGAACTGGTCACGGTAGAGACGGACGAAGAGCTTGGCGTTGCCGCGGTC
4741  ---------+---------+---------+---------+---------+---------+ 4800
      GAGGTGCACGGCCTTGACCAGTGCCATCTCTGCCTGCTTCTCGAACCGCAACGGCGCCAG
-3          E  V  H  R  F  Q  D  R  Y  L  R  V  F  L  K  A  N  G  R  D -

      GGTCAGGCTGGCGGGAATCTCGCCCGCCTCCCAGGCGGTCGCGGCGACGGGGGCCTCGGG
4801  ---------+---------+---------+---------+---------+---------+ 4860
      CCAGTCCGACCGCCCTTAGAGCGGGCGGAGGGTCCGCCAGCGCCGCTGCCCCCGGAGCCC
-3          T  L  S  A  P  I  E  G  A  E  W  A  T  A  A  V  P  A  E  P -

      AGCGGCCTGGACAGGGAGGAGCGGCGCTGGGGCCGGGGTGGTTTCGAGGGCCAGCATCTG
4861  ---------+---------+---------+---------+---------+---------+ 4920
      TCGCCGGACCTGTCCCTCCTCGCCGCGACCCCGGCCCCACCAAAGCTCCCGGTCGTAGAC
-3          A  A  Q  V  P  L  L  P  A  P  A  P  T  T  E  L  A  L  M  Q -

      CTGAGCGGCGGCAGTTGCGTCAAAGCGAGGGCCCTCGGCGCTGCTGCTCATGGACGTCCT
4921  ---------+---------+---------+---------+---------+---------+ 4980
      GACTCGCCGCCGTCAACGCAGTTTCGCTCCCGGGAGCCGCGACGACGAGTACCTGCAGGA
-3-<        Q  A  A  A  T  A  D  F  R  P  G  E  A  S  S  S  M -

      TCGAGATGGAGCGGTCGGGCGGTCCCCGCTGCGGGAACGGCATGAATGATCTTCCCGGTG
4981  ---------+---------+---------+---------+---------+---------+ 5040
      AGCTCTACCTCGCCAGCCCGCCAGGGGCGACGCCCTTGCCGTACTTACTAGAAGGGCCAC

      CGGACAGAGTGCCAGGGGCAGCGCATGTGCGGGGGGACAACGGCCCGTTTCGGACGAGGG
5041  ---------+---------+---------+---------+---------+---------+ 5100
      GCCTGTCTCACGGTCCCCGTCGCGTACACGCCCCCCTGTTGCCGGGCAAAGCCTGCTCCC

      CCGGCCGACGGGGGGAAGCAGGGGCCGGCAACCGGGTGGCGGGGCGGCGTGAGCGAGGGC
5101  ---------+---------+---------+---------+---------+---------+ 5160
      GGCCGGCTGCCCCCCTTCGTCCCCGGCCGTTGGCCCACCGCCCCGCCGCACTCGCTCCCG

      ACGAGCGGCCCGGTACGGGGGGAAGGGCTCGTCTCTCCGTGGGGCGGCACGTTGTGGTCC
5161  ---------+---------+---------+---------+---------+---------+ 5220
      TGCTCGCCGGGCCATGCCCCCCTTCCCGAGCAGAGAGGCACCCCGCCGTGCAACACCAGG

      TCGTCCGTCAGCTTGCGTCTGGCTTCAGCCTCCTGACCCCCAATAAGGCGAAAGCTGCTG
5221  ---------+---------+---------+---------+---------+---------+ 5280
      AGCAGGCAGTCGAACGCAGACCGAAGTCGGAGGACTGGGGGTTATTCCGCTTTCGACGAC

      GTCAAGCATCTTTCGTGACACTCGGCGAGGGACTGAAGGGACTGTCTTTCGGAATGAGTG
5281  ---------+---------+---------+---------+---------+---------+ 5340
      CAGTTCGTAGAAAGCACTGTGAGCCGCTCCCTGACTTCCCTGACAGAAAGCCTTACTCAC

      TAGGGGGTTGTCGGGTGGGGACCGCGCCTCGACTCCCCGGCGGACGGGATCTGTTCGGTC
5341  ---------+---------+---------+---------+---------+---------+ 5400
      ATCCCCCAACAGCCCACCCCTGGCGCGGAGCTGAGGGGCCGCCTGCCCTAGACAAGCCAG

      GGTCCCTTGGGTCCCTCCCCGGATCGCGGCAGGGACCCAAGGGGGCGGTGCGGCGGGCGG
5401  ---------+---------+---------+---------+---------+---------+ 5460
      CCAGGGAACCCAGGGAGGGGCCTAGCGCCGTCCCTGGGTTCCCCCGCCACGCCGCCCGCC

      TCGGTGAGGGGCCCCGGTGGAGGGACTGAGGGTCTGTATGGAGCGATAAGAGGGTCTGAA
5461  ---------+---------+---------+---------+---------+---------+ 5520
      AGCCACTCCCCGGGGCCACCTCCCTGACTCCCAGACATACCTCGCTATTCTCCCAGACTT

      GGGGCGGAGAGAGTTTCGGTCCCTGCGTTGAGTCCCTGGTCATCACCGCAGGTCAGAGGG
5521  ---------+---------+---------+---------+---------+---------+ 5580
      CCCCGCCTCTCTCAAAGCCAGGGACGCAACTCAGGGACCAGTAGTGGCGTCCAGTCTCCC

      GTTTTGAGGGGTGAAAAAGGGACTGAAGGGACTCAACTTCCCCATTATGAGCTGAGTAGA
5581  ---------+---------+---------+---------+---------+---------+ 5640
      CAAAACTCCCCACTTTTTTCCCTGACTTCCCTGAGTTGAAGGGGTAATACTCGACTCATCT
```

7

```
              AGAAAGCAGTATGACGATATCGGCGCCTACATACGCGCGCGTACATAGTGAGCTTATAAT
       5641   ---------+---------+---------+---------+---------+---------+  5700
              TCTTTCGTCATACTGCTATAGCCGCGGATGTATGCGCGCGCATGTATCACTCGAATATTA

              GCGGAAGTTGAGTCCCTTCAGTCCCTTTTCGTGGGGTCGTATCCCCTCTGACTGCGTTGA
       5701   ---------+---------+---------+---------+---------+---------+  5760
              CGCCTTCAACTCAGGGAAGTCAGGGAAAAGCACCCCAGCATAGGGGAGACTGACGCAACT

              CCGTCGCCGCTCCGCGCAGGGACCGAAGAGGGACCAAGTCCCTGCGCGGGGCGGGCGACG
       5761   ---------+---------+---------+---------+---------+---------+  5820
              GGCAGCGGCGAGGCGCGTCCCTGGCTTCTCCCTGGTTCAGGGACGCGCCCCGCCCGCTGC

              GTAATCGTGCAGTGCCCCCTCCCCCGTTTCCCACAGCGAGTCGTCGCTCCCCTGTGAGGC
       5821   ---------+---------+---------+---------+---------+---------+  5880
              CATTAGCACGTCACGGGGGAGGGGGCAAAGGGTGTCGCTCAGCAGCGAGGGGACACTCCG

              CGGAGAGGGTCCTAGAACCCCTCAGGGGCCGTTCTGTGGCCCTCTGGGCCTCCTCCTGGC
       5881   ---------+---------+---------+---------+---------+---------+  5940
              GCCTCTCCCAGGATCTTGGGGAGTCCCCGGCAAGACACCGGGAGACCCGGAGGAGGACCG

              CATTTACCCCATGGGGGCGCTTGGGGGCGTCAGGAGGGCTTGTGAGGGCTCTGCCGGGAA
       5941   ---------+---------+---------+---------+---------+---------+  6000
              GTAAATGGGGTACCCCCGCGAACCCCCGCAGTCCTCCCGAACACTCCCGAGACGGCCCTT
       -2->                                             M  R  A  L  P  G  S -

              GTGGCGGATTGCGCATGGCAGGAGATGCCCCGACAGCGGCCGGGAATCGACGATGTCCCC
       6001   ---------+---------+---------+---------+---------+---------+  6060
              CACCGCCTAACGCGTACCGTCCTCTACGGGGCTGTCGCCGGCCCTTAGCTGCTACAGGGG
       -2       G  G  L  R  M  A  G  D  A  P  T  A  A  G  N  R  R  C  P  P -

              CGACCCCTATCCAGCGTCCGCTGATCCTCAGGAGGCAGACCTTGCAGGCTCCAGAAGCGA
       6061   ---------+---------+---------+---------+---------+---------+  6120
              GCTGGGGATAGGTCGCAGGCGACTAGGAGTCCTCCGTCTGGAACGTCCGAGGTCTTCGCT
       -2       T  P  I  Q  R  P  L  I  L  R  R  Q  T  L  Q  A  P  E  A  K -

              AGAACGGCCGGTCCCCGGAGCAGCCGCAGGAAGAGCGGATCGTCCTGGACGTATGGCTGG
       6121   ---------+---------+---------+---------+---------+---------+  6180
              TCTTGCCGGCCAGGGGCCTCGTCGGCGTCCTTCTCGCCTAGCAGGACCTGCATACCGACC
       -2       N  G  R  S  P  E  Q  P  Q  E  E  R  I  V  L  D  V  W  L  A -

              CGAACTACCCGTTCCCCACCTATGACGGGCGTGACTTCCTCGCTCCGCTGCGCGAGCGGG
       6181   ---------+---------+---------+---------+---------+---------+  6240
              GCTTGATGGGCAAGGGGTGGATACTGCCCGCACTGAAGGAGCGAGGCGACGCGCTCGCCC
       -2       N  Y  P  F  P  T  Y  D  G  R  D  F  L  A  P  L  R  E  R  A -

              CGGCGGAGTTCGAGCGCGCCCACCCCCGATACCGGGTCGACATCAACGGCCACGACTTCT
       6241   ---------+---------+---------+---------+---------+---------+  6300
              GCCGCCTCAAGCTCGCGCGGGTGGGGGCTATGGCCCAGCTGTAGTTGCCGGTGCTGAAGA
       -2       A  E  F  E  R  A  H  P  R  Y  R  V  D  I  N  G  H  D  F  W -

              GGACCATCCCCGAGAAGGTGGCGCGCGCCACCGCGGAGGGCAGGCCTCCGCACATAGCGG
       6301   ---------+---------+---------+---------+---------+---------+  6360
              CCTGGTAGGGGCTCTTCCACCGCGCGCGGTGGCGCCTCCCGTCCGGAGGCGTGTATCGCC
       -2       T  I  P  E  K  V  A  R  A  T  A  E  G  R  P  P  H  I  A  G -

              GCTACTACGCCACCGACAGCCAGTTGGCGCGGGACGCGCGCAGGCCCGACGGGAAGCCGG
       6361   ---------+---------+---------+---------+---------+---------+  6420
              CGATGATGCGGTGGCTGTCGGTCAACCGCGCCCTGCGCGCGTCCGGGCTGCCCTTCGGCC
       -2       Y  Y  A  T  D  S  Q  L  A  R  D  A  R  R  P  D  G  K  P  V -

              TCTTCACCTCGGTGGAGGCCGCGTTGGCCGGCCGGACGGAGATACTGGGACACCCGGTGG
       6421   ---------+---------+---------+---------+---------+---------+  6480
              AGAAGTGGAGCCACCTCCGGCGCAACCGGCCGGCCTGCCTCTATGACCCTGTGGGCCACC
       -2       F  T  S  V  E  A  A  L  A  G  R  T  E  I  L  G  H  P  V  V -

              TGGTGGAGGACCTCGACCCCGTGGTGCGCGACTCCTACTCGTTCGGGGGCGAGTTGGTGT
```

8

```
6481  ---------+---------+---------+---------+---------+---------+  6540
      ACCACCTCCTGGAGCTGGGGCACCACGCGCTGAGGATGAGCAAGCCCCCGCTCAACCACA
-2     V  E  D  L  D  P  V  V  R  D  S  Y  S  F  G  G  E  L  V  S -

      CGCTGCCGCTCACGGTCACCACCATGCTCTGCTACGCCAACTCCTCCCTCCTCGCGCGCG
6541  ---------+---------+---------+---------+---------+---------+  6600
      GCGACGGCGAGTGCCAGTGGTGGTACGAGACGATGCGGTTGAGGAGGGAGGAGCGCGCGC
-2     L  P  L  T  V  T  T  M  L  C  Y  A  N  S  S  L  L  A  R  A -

      CCGGTGTTCCGGAGTTGCCCCGTACCTGGGATGAGGTCGAAGCAGCCTGCCAGGCGGTGG
6601  ---------+---------+---------+---------+---------+---------+  6660
      GGCCACAAGGCCTCAACGGGGCATGGACCCTACTCCAGCTTCGTCGGACGGTCCGCCACC
-2     G  V  P  E  L  P  R  T  W  D  E  V  E  A  A  C  Q  A  V  A -

      CCAGCGTCGACGGGGGGCCCGGTCACGGAATCACCTGGGCCAACGACGGCTGGGTTTTCC
6661  ---------+---------+---------+---------+---------+---------+  6720
      GGTCGCAGCTGCCCCCCGGGCCAGTGCCTTAGTGGACCCGGTTGCTGCCGACCCAAAAGG
-2     S  V  D  G  G  P  G  H  G  I  T  W  A  N  D  G  W  V  F  Q -

      AGCAGGCCGTCGCCCTTCAGAACGGGGTGCTGACCGATCAGGACAACGGCCGCTCCGGCT
6721  ---------+---------+---------+---------+---------+---------+  6780
      TCGTCCGGCAGCGGGAAGTCTTGCCCCACGACTGGCTAGTCCTGTTGCCGGCGAGGCCGA
-2     Q  A  V  A  L  Q  N  G  V  L  T  D  Q  D  N  G  R  S  G  S -

      CCGCCACGACGGTGGACGTCACATCGGACGAGATGCTGGACTGGGTCCGCTGGTGGACGC
6781  ---------+---------+---------+---------+---------+---------+  6840
      GGCGGTGCTGCCACCTGCAGTGTAGCCTGCTCTACGACCTGACCCAGGCGACCACCTGCG
-2     A  T  T  V  D  V  T  S  D  E  M  L  D  W  V  R  W  W  T  H -

      ACCTCCATGAGCGCGGCCATTACCTCTACACGGGCGGGCCCTCGGACTGGGGCGGGGCGT
6841  ---------+---------+---------+---------+---------+---------+  6900
      TGGAGGTACTCGCGCCGGTAATGGAGATGTGCCCGCCCGGGAGCCTGACCCCGCCCCGCA
-2     L  H  E  R  G  H  Y  L  Y  T  G  G  P  S  D  W  G  A  F -

      TCGAGGCTTTCGTCCAGCAGAAGGTCGCATTCACCTTCGACTCGTCCAAGGCCGCCCGGG
6901  ---------+---------+---------+---------+---------+---------+  6960
      AGCTCCGAAAGCAGGTCGTCTTCCAGCGTAAGTGGAAGCTGAGCAGGTTCCGGCGGGCCC
-2     E  A  F  V  Q  Q  K  V  A  F  T  F  D  S  S  K  A  A  R  E -

      AACTCATCCAGGCCGGTGCACAGGCCGGTTTCGAGGTCGCGGTGTTCCCGTTGCCCAGGA
6961  ---------+---------+---------+---------+---------+---------+  7020
      TTGAGTAGGTCCGGCCACGTGTCCGGCCAAAGCTCCAGCGCCACAAGGGCAACGGGTCCT
-2     L  I  Q  A  G  A  Q  A  G  F  E  V  A  V  F  P  L  P  R  N -

      ACGCGAAGGCCCCGGTAGCGGGCCAGCCCGTCTCGGGAGACTCCCTGTGGCTGGCCGCGG
7021  ---------+---------+---------+---------+---------+---------+  7080
      TGCGCTTCCGGGGCCATCGCCCGGTCGGGCAGAGCCCTCTGAGGGACACCGACCGGCGCC
-2     A  K  A  P  V  A  G  Q  P  V  S  G  D  S  L  W  L  A  A  G -

      GACTCGACGAGACCACGCAGGACGGGCTGCTCGCTCTCACCCAGTACCTGATCAGCCCGG
7081  ---------+---------+---------+---------+---------+---------+  7140
      CTGAGCTGCTCTGGTGCGTCCTGCCCGACGAGCGAGAGTGGGTCATGGACTAGTCGGGCC
-2     L  D  E  T  T  Q  D  G  L  L  A  L  T  Q  Y  L  I  S  P  A -

      CCAACGCCGCGGACTGGCACCGCACCAACGGTTTCGTACCGGTGACCGGCGCGGCCGGGG
7141  ---------+---------+---------+---------+---------+---------+  7200
      GGTTGCGGCGCCTGACCGTGGCGTGGTTGCCAAAGCATGGCCACTGGCCGCGCCGGCCCC
-2     N  A  A  D  W  H  R  T  N  G  F  V  P  V  T  G  A  A  G  E -

      AACTGCTGGAAGCGACAGGCTGGTTCGACCGCCGGCCGCAGCAACGGGTGGCCGGGGAGC
7201  ---------+---------+---------+---------+---------+---------+  7260
      TTGACGACCTTCGCTGTCCGACCAAGCTGGCGGCCGGCGTCGTTGCCCACCGGCCCCTCG
-2     L  L  E  A  T  G  W  F  D  R  R  P  Q  Q  R  V  A  G  E  Q -

      AGTTGAAGGCGTCCGACCGGTCACCGGCGGCGCTCGGCGCGCTGCTCGGCGACTTCGCGG
7261  ---------+---------+---------+---------+---------+---------+  7320
```

9

```
                         TCAACTTCCGCAGGCTGGCCAGTGGCCGCCGCGAGCCGCGCGACGAGCCGCTGAAGCGCC
         -2               L  K  A  S  D  R  S  P  A  A  L  G  A  L  L  G  D  F  A  A  -

                         CCGTCAACGAGGTCATCACCGCAGCGATGGACGATGTCCTGCGCAGTGGAGCGGACCCCG
         7321  ---------+---------+---------+---------+---------+---------+ 7380
                         GGCAGTTGCTCCAGTAGTGGCGTCGCTACCTGCTACAGGACGCGTCACCTCGCCTGGGGC
         -2               V  N  E  V  I  T  A  A  M  D  D  V  L  R  S  G  A  D  P  A  -

                         CGAAGGCCTTCGCCGAAGCCGGCGTGGCCGCCCAGCAACTGCTCGATGCCTACAACGCCC
         7381  ---------+---------+---------+---------+---------+---------+ 7440
                         GCTTCCGGAAGCGGCTTCGGCCGCACCGGCGGGTCGTTGACGAGCTACGGATGTTGCGGG
         -2               K  A  F  A  E  A  G  V  A  A  Q  Q  L  L  D  A  Y  N  A  R  -

                         GGAACCGCTCCGGATCCGGGACCCCCTCCGCCGTCTGAGATCCGGTACCGGGGCACAGGG
         7441  ---------+---------+---------+---------+---------+---------+ 7500
                         CCTTGGCGAGGCCTAGGCCCTGGGGGAGGCGGCAGACTCTAGGCCATGGCCCCGTGTCCC
         -2-*             N  R  S  G  S  G  T  P  S  A  V  *  -

                         GCGCCGCCGCCCGCTTTCCCGGCGGGGCACTGGCCGGGGGACATGCTCTCCCGCCCCCGG
         7501  ---------+---------+---------+---------+---------+---------+ 7560
                         CGCGGCGGCGGGCGAAAGGGCCGCCCCGTGACCGGCCCCCTGTACGAGAGGGCGGGGGCC

                         CAGGACGTAGGGTCAACCCGCCTGCGCCTTCAGGTGGCGGCGCAGATACTCACCGGTCAG
         7561  ---------+---------+---------+---------+---------+---------+ 7620
                         GTCCTGCATCCCAGTTGGGCGGACGCGGAAGTCCACCGCCGCGTCTATGAGTGGCCAGTC
         -1-*                  *  G  A  Q  A  K  L  H  R  R  L  Y  E  G  T  L  -

                         GGAGGAATCCGCGGCGAGCAGGTCCTTCGGTGTGCCGGTGAAGACGATCTCGCCGCCCTC
         7621  ---------+---------+---------+---------+---------+---------+ 7680
                         CCTCCTTAGGCGCCGCTCGTCCAGGAAGCCACACGGCCACTTCTGCTAGAGCGGCGGGAG
         -1               S  S  D  A  A  L  L  D  K  P  T  G  T  F  V  I  E  G  G  E  -

                         CCGTCCCCCGTCGGGACCCAGGTCGATGATCCAGTCGGCCTGCTGCACCACATCGAGGTT
         7681  ---------+---------+---------+---------+---------+---------+ 7740
                         GGCAGGGGGCAGCCCTGGGTCCAGCTACTAGGTCAGCCGGACGACGTGGTGTAGCTCCAA
         -1               R  G  G  D  P  G  L  D  I  I  W  D  A  Q  Q  V  V  D  L  N  -

                         GTGCTCGATGACCACGACGGTGTTCCCGGCCTCGACGAGCCCGTCCAGGAGCTTCAGCAG
         7741  ---------+---------+---------+---------+---------+---------+ 7800
                         CACGAGCTACTGGTGCTGCCACAAGGGCCGGAGCTGCTCGGGCAGGTCCTCGAAGTCGTC
         -1               H  E  I  V  V  V  T  N  G  A  E  V  L  G  D  L  L  K  L  L  -

                         GGTGTCAACGTCCGACATGTGCAGCCCGGTGGTGGGCTCGTCCAGGACATAGACCGTGCC
         7801  ---------+---------+---------+---------+---------+---------+ 7860
                         CCACAGTTGCAGGCTGTACACGTCGGGCCACCACCCGAGCAGGTCCTGTATCTGGCACGG
         -1               T  D  V  D  S  M  H  L  G  T  T  P  E  D  L  V  Y  V  T  G  -

                         CGTGCGGTGCAGCTGGTCGGCAAGTTTGATCCGCTGCAGTTCACCGCCGGAGAGGCTGGA
         7861  ---------+---------+---------+---------+---------+---------+ 7920
                         GCACGCCACGTCGACCAGCCGTTCAAACTAGGCGACGTCAAGTGGCGGCCTCTCCGACCT
         -1               T  R  H  L  Q  D  A  L  K  I  R  Q  L  E  G  G  S  L  S  S  -

                         AAGCGGCTGGCCCAGGCTGAGGTACCCAAGACCGACGTCGACGAGAGCGCGCAGTTTCGG
         7921  ---------+---------+---------+---------+---------+---------+ 7980
                         TTCGCCGACCGGGTCCGACTCCATGGGTTCTGGCTGCAGCTGCTCTCGCGCGTCAAAGCC
         -1               L  P  Q  G  L  S  L  Y  G  L  G  V  D  V  L  A  R  L  K  P  -

                         CAGCAGGGCCTTCTCGGTGAAGAACTCGACGGCCTCGTCGGCGGGCAGCTCCAGGACGTC
         7981  ---------+---------+---------+---------+---------+---------+ 8040
                         GTCGTCCCGGAAGAGCCACTTCTTGAGCTGCCGGAGCAGCCGCCCGTCGAGGTCCTGCAG
         -1               L  L  A  K  E  T  F  F  E  V  A  E  D  A  P  L  E  L  V  D  -

                         CGCGATCGACTTCCCGCGAAGCTGGTGCTCCAGGACCTCGGGCTTGAAGCGGCGCCCCTC
         8041  ---------+---------+---------+---------+---------+---------+ 8100
                         GCGCTAGCTGAAGGGCGCTTCGACCACGAGGTCCTGGAGCCCGAACTTCGCCGCGGGGAG
         -1               A  I  S  K  G  R  L  Q  H  E  L  V  E  P  K  F  R  R  G  E  -
```

10

```
                ACAGACACCGCAGTGCGTGGTCACCGGATCCATGAAGGCCAGCTCGGTGATGATGACCCC
         8101   ---------+---------+---------+---------+---------+---------+ 8160
                TGTCTGTGGCGTCACGCACCAGTGGCCTAGGTACTTCCGGTCGAGCCACTACTACTGGGG
    -1            C  V  G  H  T  T  V  P  D  M  F  A  L  E  T  I  I  V  G -

                GCGGCCCTGGCACTCCTCGCACGACCCCTTGGAGTTGAAGCTGAACAGCGAGGCGTTCGC
         8161   ---------+---------+---------+---------+---------+---------+ 8220
                CGCCGGGACCGTGAGGAGCGTGCTGGGGAACCTCAACTTCGACTTGTCGCTCCGCAAGCG
    -1            R  G  Q  C  E  E  C  S  G  K  S  N  F  S  F  L  S  A  N  A -

                GCCGGTCTCCTTCGCGAACAGCTTGCGCAGCGGGTCCATCAGGCCGAGGTAGGAGACCGG
         8221   ---------+---------+---------+---------+---------+---------+ 8280
                CGGCCAGAGGAAGCGCTTGTCGAACGCGTCGCCCAGGTAGTCCGGCTCCATCCTCTGGCC
    -1            G  T  E  K  A  F  L  K  R  L  P  D  M  L  G  L  Y  S  V  P -

                TGTGGAGCGCGACGAGGCGGCGATCGCGGACTGGTCGACAAAGACCGCGTCGGGGTGCGC
         8281   ---------+---------+---------+---------+---------+---------+ 8340
                ACACCTCGCGCTGCTCCGCCGCTAGCGCCTGACCAGCTGTTTCTGGCGCAGCCCCACGCG
    -1            T  S  R  S  S  A  A  I  A  S  Q  D  V  F  V  A  D  P  H  A -

                CTCCATGAATGCCCCGGAGATCAGGCTGCTCTTGCCGGAACCCGCCACCCCGGTCACCGC
         8341   ---------+---------+---------+---------+---------+---------+ 8400
                GAGGTACTTACGGGGCCTCTAGTCCGACGAGAACGGCCTTGGGCGGTGGGGCCAGTGGCG
    -1            E  M  F  A  G  S  I  L  S  S  K  G  S  G  A  V  G  T  V  A -

                GGTCAGCACACCGGTGGGCACGGCCACGGAGACCTGCTTCAGGTTGTGGAGATCCGCGTT
         8401   ---------+---------+---------+---------+---------+---------+ 8460
                CCAGTCGTGTGGCCACCCGTGCCGGTGCCTCTGGACGAAGTCCAACACCTCTAGGCGCAA
    -1            T  L  V  G  T  P  V  A  V  S  V  Q  K  L  N  H  L  D  A  N -

                CTCCACGGTCAGCTCCCCCGTGGGCGGGCGGACCTCCTCCTTCACGCGGGCCCCCCGCCG
         8461   ---------+---------+---------+---------+---------+---------+ 8520
                GAGGTGCCAGTCGAGGGGGCACCCGCCCGCCTGGAGGAGGAAGTGCGCCCGGGGGGCGGC
    -1            E  V  T  L  E  G  T  P  P  R  V  E  E  K  V  R  A  G  R  R -

                CAGAGCCTCCCCGGTCCGGGTCTTCGCCTTCCGCAGCTTCGCGAAGGACCCCTCGAACAC
         8521   ---------+---------+---------+---------+---------+---------+ 8580
                GTCTCGGAGGGGCCAGGCCCAGAAGCGGAAGGCGTCGAAGCGCTTCCTGGGGAGCTTGTG
    -1            L  A  E  G  T  R  T  K  A  K  R  L  K  A  F  S  G  E  F  V -

                GATCTCGCCCCCGTGCACTCCCGCCCCGGGACCGACATCGACGATGTGGTCGGCGATCTC
         8581   ---------+---------+---------+---------+---------+---------+ 8640
                CTAGAGCGGGGGCACGTGAGGGCGGGGCCCTGGCTGTAGCTGCTACACCAGCCGCTAGAG
    -1            I  E  G  G  H  V  G  A  G  P  G  V  D  V  I  H  D  A  I  E -

                GATCACAtcGGGGTCGTGCTCGACGACCAGCACGGTGTTCCCCTTGTCGCGCAGCGCGCG
         8641   ---------+---------+---------+---------+---------+---------+ 8700
                CTAGTGTagCCCCAGCACGAGCTGCTGGTCGTGCCACAAGGGGAACAGCGCGTCGCGCGC
    -1            I  V  D  P  D  H  E  V  V  L  V  T  N  G  K  D  R  L  A  R -

                CAGCAGGTCGTTGAGCCGCCCCACGTCGCGCGGGTGCAGGCCGATGCTGGGCTCGTCGAA
         8701   ---------+---------+---------+---------+---------+---------+ 8760
                GTCGTCCAGCAACTCGGCGGGGTGCAGCGCGCCCACGTCCGGCTACGACCCGAGCAGCTT
    -1            L  L  D  N  L  R  G  V  D  R  P  H  L  G  I  S  P  E  D  F -

                GATGTACGTGAGCCCGGCCAGACCACTGCCGAGGTGGCGCACCATCTTCAGCCGCTGCCC
         8761   ---------+---------+---------+---------+---------+---------+ 8820
                CTACATGCACTCGGGCCGGTCTGGTGACGGCTCCACCGCGTGGTAGAAGTCGGCGACGGG
    -1            I  Y  T  L  G  A  L  G  S  G  L  H  R  V  M  K  L  R  Q  G -

                CTCGCCCCCCGAGAGGTCGGCCGTGGGCCTGTCCAGGGTCAGGTAGCCGAGCCCGATGGA
         8821   ---------+---------+---------+---------+---------+---------+ 8880
                GAGCGGGGGGCTCTCCAGCCGGCACCCGGACAGGTCCCAGTCCATCGGCTCGGGCTACCT
    -1            E  G  G  S  L  D  A  T  P  R  D  L  T  L  Y  G  L  G  I  S -
```

11

```
        CACGATCCGCTCCAGGGCCGTGCGCGCGGCTTTCGCGAGAGGGGCAGCGGCCGGCTCCGT
8881    ---------+---------+---------+---------+---------+---------+ 8940
        GTGCTAGGCGAGGTCCCGGCACGCGCGCCGAAAGCGCTCTCCCCGTCGCCGGCCGAGGCA
-1        V  I  R  E  L  A  T  R  A  A  K  A  L  P  A  A  A  P  E  T  -

        GACGCCGGCGAGCACCTCCGTGAGGTCGCGGACCTCCATGCTCGAGTAGTCGGCGATGTT
8941    ---------+---------+---------+---------+---------+---------+ 9000
        CTGCGGCCGCTCGTGGAGGCACTCCAGCGCCTGGAGGTACGAGCTCATCAGCCGCTACAA
-1        V  G  A  L  V  E  T  L  D  R  V  E  M  S  S  Y  D  A  I  N  -

        CTTGCCGTCGATCCGGACGTCGAGCGCGGCGGCGTTGAGCCGCGCGCCCCGGCAGGAGGG
9001    ---------+---------+---------+---------+---------+---------+ 9060
        GAACGGCAGCTAGGCCTGCAGCTCGCGCCGCCGCAACTCGGCGCGCGGGGCCGTCCTCCC
-1        K  G  D  I  R  V  D  L  A  A  A  N  L  R  A  G  R  C  S  P  -

        ACAGACTCCGTCgGTGACGAAACGTTCGATGACCTCGCGCTTGCGGTCGcTCAGCGCGCT
9061    ---------+---------+---------+---------+---------+---------+ 9120
        TGTCTGAGGCAGcCACTGCTTTGCAAGCTACTGGAGCGCGAACGCCAGCgAGTCGCGCGA
-1        C  V  G  D  T  V  F  R  E  I  V  E  R  K  R  D  S  L  A  S  -

        GAGGTCGCGCTTGAGGTTGAgCCGCTCGAACCGGTCGGCcAACCCCTCGTAGTTCGTCTG
9121    ---------+---------+---------+---------+---------+---------+ 9180
        CTCCAGCGCGAACTCCAACTcGGCGAGCTTGGCCAGCCGgTTGGGGAGCATCAAGCAGAC
-1        L  D  R  K  L  N  L  R  E  F  R  D  A  L  G  E  Y  N  T  Q  -

        GAACTCGGTGCTCTTGGTCTTCAGCGTcACCTTCCCGCCGGTGcCGCGCAGCAGCGTGTC
9181    ---------+---------+---------+---------+---------+---------+ 9240
        CTTGAGCCACGAGAACCAGAAGTCGCAgTGGAAGGGCGGCCACgGCGCGTCGTCGCACAG
-1        F  E  T  S  K  T  K  L  T  V  K  G  G  T  G  R  L  L  T  D  -

        CAGCTCCTCGGCGCTGTACTCGGCGATCGGCTTGGCCGGATCCAGACGGCCGGACTTCGC
9241    ---------+---------+---------+---------+---------+---------+ 9300
        GTCGAGGAGCCGCGACATGAGCCGCTAGCCGAACCGGCCTAGGTCTGCCGGCCTGAAGCG
-1        L  E  E  A  S  Y  E  A  I  P  K  A  P  D  L  R  G  S  K  A  -

        CCAGATCTGCCAGTCCGGGCTACCCACCTTGTACTCGGGGAAAAGGACCGCCCCGTCGTC
9301    ---------+---------+---------+---------+---------+---------+ 9360
        GGTCTAGACGGTCAGGCCCGATGGGTGGAACATGAGCCCCTTTTCCTGGCGGGGCAGCAG
-1        W  I  Q  W  D  P  S  G  V  K  Y  E  P  F  L  V  A  G  D  D  -

        CAGGGACTTCGAGCGGTCCAGCATCTTGTCCAGGTCGAGGGCGATGCTCTGGCCGAGACC
9361    ---------+---------+---------+---------+---------+---------+ 9420
        GTCCCTGAAGCTCGCCAGGTCGTAGAACAGGTCCAGCTCCCGCTACGAGACCGGCTCTGG
-1        L  S  K  S  R  D  L  M  K  D  L  D  L  A  I  S  Q  G  L  G  -

        GTCGCAGTCCGGGCACATGCCCTGGGGGTCGTTGAACGAGAACGCGGAGACGCCGAGCGA
9421    ---------+---------+---------+---------+---------+---------+ 9480
        CAGCGTCAGGCCCGTGTACGGGACCCCCAGCAACTTGCTCTTGCGCCTCTGCGGCTCGCT
-1        D  C  D  P  C  M  G  Q  P  D  N  F  S  F  A  S  V  G  L  S  -

        GGACGGCCCGTCGTCCTTCGTCGTGCCGAACCGTGCGAACAGGGCCCGGATCATCGGCTG
9481    ---------+---------+---------+---------+---------+---------+ 9540
        CCTGCCGGGCAGCAGGAAGCAGCACGGCTTGGCACGCTTGTCCCGGGCCTAGTAGCCGAC
-1        S  P  G  D  D  K  T  T  G  F  R  A  F  L  A  R  I  M  P  Q  -

        TACGTCCGTCATGGTCCCCACCGTGGACCGGGCGTTGCCCCCCACGGGCTTCTGGTCGAC
9541    ---------+---------+---------+---------+---------+---------+ 9600
        ATGCAGGCAGTACCAGGGGTGGCACCTGGCCCGCAACGGGGGGTGCCCGAAGACCAGCTG
-1        V  D  T  M  T  G  V  T  S  R  A  N  G  G  V  P  K  Q  D  V  -

        GATCACCGGGGTGGTGAGGTTCTCGATCGCCTCGGCCTGAGGACGTTCGTACTTCGGAAG
9601    ---------+---------+---------+---------+---------+---------+ 9660
        CTAGTGGCCCCACCACTCCAAGAGCTAGCGGAGCCGGACTCCTGCAAGCATGAAGCCTTC
-1        I  V  P  T  T  L  N  E  I  A  E  A  Q  P  R  E  Y  K  P  L  -

        CTGGTTGCGGATGTACCAGCTGAAGGTGGAGTTCAGCTGTCGCTGGGCCTCCACGGCCAC
```

12

```
                               ●                           ●

              GACCAACGCCTACATGGTCGACTTCCACCTCAAGTCGACAGCGACCCGGAGGTGCCGGTG
       9661   ---------+---------+---------+---------+---------+---------+   9720
              CTGGTTGCGGATGTACCAGCTGAAGGTGGAGTTCAGCTGTCGCTGGGCCTCCACGGCCAC
       -1         Q   N   R   I   Y   W   S   F   T   S   N   L   Q   R   Q   A   E   V   A   V   -

              CGTGTCGAAGACGATCGACGACTTGCCCGAACCCGAGACCCCCGTGAAGACCGTGATCTG
       9721   ---------+---------+---------+---------+---------+---------+   9780
              GCACAGCTTCTGCTAGCTGCTGAACGGGCTTGGGCTCTGGGGGCACTTCTGGCACTAGAC
       -1         T   D   F   V   I   S   S   K   G   S   G   S   V   G   T   F   V   T   I   Q   -

              GTTGCGGGGAATCGTCAGGGAGACATCTTTGAGGTTGTGGATCCGCGCGCCCGCGATGCG
       9781   ---------+---------+---------+---------+---------+---------+   9840
              CAACGCCCCTTAGCAGTCCCTCTGTAGAAACTCCAACACCTAGGCGCGCGGGCGCTACGC
       -1         N   R   P   I   T   L   S   V   D   K   L   N   H   I   R   A   G   A   I   R   -

              GATGCCGTCTCCCGGGCCGGATGTTTTTCCCGCGCCGGCGGTGGGGTCGGTGACGCTCAC
       9841   ---------+---------+---------+---------+---------+---------+   9900
              CTACGGCAGAGGGCCCGGCCTACAAAAAGGGCGCGGCCGCCACCCCAGCCACTGCGAGTG
       -1-<       I   G   D   G   P   G   S   T   K   G   A   G   A   T   P   D   T   V   S   M-

              AGAGTTTTCCTCCTGGCTTCCGTACATGATTTACCGTGTCAGCCGGGCAAACCGGCGGAA
       9901   ---------+---------+---------+---------+---------+---------+   9960
              TCTCAAAAGGAGGACCGAAGGCATGTACTAAATGGCACAGTCGGCCCGTTTGGCCGCCTT

              CGGTAACCACCTAGCTTGTACTCAGGAGGTGTCCGGGGTCTTCTCCTCCCGTGCTGACTT
       9961   ---------+---------+---------+---------+---------+---------+   10020
              GCCATTGGTGGATCGAACATGAGTCCTCCACAGGCCCCAGAAGAGGAGGGCACGACTGAA
       0-*                                *   S   T   D   P   T   K   E   E   R   A   S   K   -

              GGGGGCCGGCCCGCCGGACAGGGCCGGCTCCGTGTTCCACCCCGCCAGCCGATCCCCCCG
       10021  ---------+---------+---------+---------+---------+---------+   10080
              CCCCCGGCCGGGCGGCCTGTCCCGGCCGAGGCACAAGGTGGGGCGGTCGGCTAGGGGGGC
       0          P   A   P   G   G   S   L   A   P   E   T   N   W   G   A   L   R   D   G   R   -

              CTCCGTCTCGTCCTCCTCGAGAACGATCCGGCTGCTCGCCCAGCGCAGGATCGGCGGCGC
       10081  ---------+---------+---------+---------+---------+---------+   10140
              GAGGCAGAGCAGGAGGAGCTCTTGCTAGGCCGACGAGCGGGTCGCGTCCTAGCCGCCGCG
       0          E   T   E   D   E   E   L   V   I   R   S   S   A   W   R   L   I   P   P   A   -

              CGTCACCGAGGTGATGAGGGCGACCAGCACGATGATCGTGAAGGTCACGGTGTCCAGTAC
       10141  ---------+---------+---------+---------+---------+---------+   10200
              GCAGTGGCTCCACTACTCCCGCTGGTCGTGCTACTAGCACTTCCAGTGCCACAGGTCATG
       0          T   V   S   T   I   L   A   V   L   V   I   I   T   F   T   V   T   D   L   V   -

              GCCGATACGCAGGCCGACCAGGGCGATCACCACCTCGATCATTCCACGCGAGTTCATCCC
       10201  ---------+---------+---------+---------+---------+---------+   10260
              CGGCTATGCGTCCGGCTGGTCCCGCTAGTGGTGGAGCTAGTAAGGTGCGCTCAAGTAGGG
       0          G   I   R   L   G   V   L   A   I   V   V   E   I   M   G   R   S   N   M   G   -

              CGCTCCGAGCGCCAGCCCCTCGTAGCGGCTCATCCCGCCACTACGGGCGGCGACGTACGC
       10261  ---------+---------+---------+---------+---------+---------+   10320
              GCGAGGCTCGCGGTCGGGGAGCATCGCCGAGTAGGGCGGTGATGCCCGCCGCTGCATGCG
       0          A   G   L   A   L   G   E   Y   R   S   M   G   G   S   R   A   A   V   Y   A   -

              ACCGGCGAACTTGCCGAAAGTGGCCACCAACAGCACCCCGAGGCCCGTGAGCAGCACCGA
       10321  ---------+---------+---------+---------+---------+---------+   10380
              TGGCCGCTTGAACGGCTTTCACCGGTGGTTGTCGTGGGGCTCCGGGCACTCGTCGTGGCT
       0          G   A   F   K   G   F   T   A   V   L   L   V   G   L   G   T   L   L   V   S   -

              CGGCTCCGCGAGTGCGGTCAGGTCCATGCGAAGCCCCACACTGCCCAGGAACACCGGTGC
       10381  ---------+---------+---------+---------+---------+---------+   10440
              GCCGAGGCGCTCACGCCAGTCCAGGTACGCTTCGGGGTGTGACGGGTCCTTGTGGCCACG
       0          P   E   A   L   A   T   L   D   M   R   L   G   V   S   G   L   F   V   P   A   -

              GAACACGGCCATGACCAGCGTGCGCAGCGGGGCGAGCCGTACCGGGGCGATGTGCCTCAG
       10441  ---------+---------+---------+---------+---------+---------+   10500
              CTTGTGCCGGTACTGGTCGCACGCGTCGCCCCGCTCGGCATGGCCCCGCTACACGGAGTC
```

13

```
                      F   V   A   M   V   L   T   R   L   P   A   L   R   V   P   A   I   H   R   L  -
        0
               CAGGGTCGCACCGGCCACGAACGCCCCGAACAACGCCTCCATCCCGGCCGCCGCGGTCAG
       10501   ---------+---------+---------+---------+---------+---------+  10560
               GTCCCAGCGTGGCCGGTGCTTGCGGGGCTTGTTGCGGAGGTAGGGCCGGCGGCGCCAGTC
        0       L   T   A   G   A   V   F   A   G   F   L   A   E   M   G   A   A   A   T   L  -

               CGCCCCGTACAGGACGACCACGGCCACGCCGACGGTGACGGCCGATACGGGGACCCGGCT
       10561   ---------+---------+---------+---------+---------+---------+  10620
               GCGGGGCATGTCCTGCTGGTGCCGGTGCGGCTGCCACTGCCGGCTATGCCCCTGGGCCGA
        0       A   G   Y   L   V   V   V   A   V   G   V   T   V   A   S   V   P   V   R   S  -

               GTCACCCGTACGGGACAGCCGCCTGCCGATCGGGCCGCCCACCGCACACGCCGCGGCGAC
       10621   ---------+---------+---------+---------+---------+---------+  10680
               CAGTGGGCATGCCCTGTCGGCGGACGGCTAGCCCGGCGGGTGGCGTGTGCGGCGCCGCTG
        0       D   G   T   R   S   L   R   R   G   I   P   G   G   V   A   C   A   A   A   V  -

               GAAGACGGTCGTCCAGGCCATCGTGGTCAGGACCACGGGCCCCCCGGCCGCCCCACTCGC
       10681   ---------+---------+---------+---------+---------+---------+  10740
               CTTCTGCCAGCAGGTCCGGTAGCACCAGTCCTGGTGCCCGGGGGGCCGGCGGGGTGAGCG
        0       F   V   T   T   W   A   M   T   T   L   V   V   P   G   A   A   G   S   A  -

               CAGCGCCGTCACCAGAGCGAGCAGCAGCCAGCCCACCGCGTCGTCGAACACCGCTGCCGC
       10741   ---------+---------+---------+---------+---------+---------+  10800
               GTCGCGGCAGTGGTCTCGCTCGTCGTCGGTCGGGTGGCGCAGCAGCTTGTGGCGACGGCG
        0       L   A   T   V   L   A   L   L   W   G   V   A   D   D   F   V   A   A   A  -

               GATGAGCAGCTGGCCGACGTTGCGGTGCGTCAGATTCAGGTCGGCGAGCGTCTTGGCGAT
       10801   ---------+---------+---------+---------+---------+---------+  10860
               CTACTCGTCGACCGGCTGCAACGCCACGCAGTCTAAGTCCAGCCGCTCGCAGAACCGCTA
        0       I   L   L   Q   G   V   N   R   H   T   L   N   L   D   A   L   T   K   A   I  -

               CACCGGGAGGGCCGTGACACACATCGCGACCCCGAGGAACAGCGCGAAGACGCCCCGCTC
       10861   ---------+---------+---------+---------+---------+---------+  10920
               GTGGCCCTCCCGGCACTGTGTGTAGCGCTGGGGCTCCTTGTCGCGCTTCTGCGGGGCGAG
        0       V   P   L   A   T   V   C   M   A   V   G   L   F   L   A   F   V   G   R   E  -

               TCCGGAGTCCGCGAGCAGCGAGGCGGGCACCAGGTAGCCGGTGGCGATGCCCAGCCCCAG
       10921   ---------+---------+---------+---------+---------+---------+  10980
               AGGCCTCAGGCGCTCGTCGCTCCGCCCGTGGTCCATCGGCCACCGCTACGGGTCGGGGTC
        0       G   S   D   A   L   L   S   A   P   V   L   Y   G   T   A   I   G   L   G   L  -

               AGGAATCAGAAGACCCGCCAGGCTGACCCGGGCGGCCAGACCCCCGCGCTTGCGCAGGAT
       10981   ---------+---------+---------+---------+---------+---------+  11040
               TCCTTAGTCTTCTGGGCGGTCCGACTGGGCCCGCCGGTCTGGGGGCGCGAACGCGTCCTA
        0       P   I   L   L   G   A   L   S   V   R   A   A   L   G   G   R   K   R   L   I  -

               CCGGGGGTCGAACTGGGCACCTGCGATGGCCACCAGCAGAAGGACGCCGAACTGGCAGAA
       11041   ---------+---------+---------+---------+---------+---------+  11100
               GGCCCCCAGCTTGACCCGTGGACGCTACCGGTGGTCGTCTTCCTGCGGCTTGACCGTCTT
        0       R   P   D   F   Q   A   G   A   I   A   V   L   L   L   V   G   F   Q   C   F  -

               CGCGTCGAGCAGGTGCGCCTGCGAGATGTCCTCGGGAAACAGCCTGCCGGAAAGTCCCGG
       11101   ---------+---------+---------+---------+---------+---------+  11160
               GCGCAGCTCGTCCACGCGGACGCTCTACAGGAGCCCTTTGTCGGACGGCCTTTCAGGGCC
        0       A   D   L   L   H   A   Q   S   I   D   E   P   F   L   R   G   S   L   G   P  -

               CGAGATCTGCCCCAGCAGGGTCGGCCCGAGCAGTACCCCCGCGGTCAGCTCCCCCACCAG
       11161   ---------+---------+---------+---------+---------+---------+  11220
               GCTCTAGACGGGGTCGTCCCAGCCGGGCTCGTCATGGGGGCGCCAGTCGAGGGGGTGGTC
        0       S   I   Q   G   L   L   T   P   G   L   L   V   G   A   T   L   E   G   V   L  -

               CGGCGGCAGACCGATCCGGGTCCCCAGCCGTCCCAGACCGTAGGCACAGGCGAGCAGGAG
       11221   ---------+---------+---------+---------+---------+---------+  11280
               GCCGCCGTCTGGCTAGGCCCAGGGGTCGGCAGGGTCTGGCATCCGTGTCCGCTCGTCCTC
        0       P   P   L   G   I   R   T   G   L   R   G   L   G   Y   A   C   A   L   L   L  -
```

14

```
          GCCGACCTGGAGCAGGAAGACCGTCAGCGGCTCCCCGCCCAGCGGCGACGTGGCTGCGAG
    11281 ---------+---------+---------+---------+---------+---------+ 11340
          CGGCTGGACCTCGTCCTTCTGGCAGTCGCCGAGGGGCGGGTCGCCGCTGCACCGACGCTC
    0        G  V  Q  L  L  F  V  T  L  P  E  G  G  L  P  S  T  A  A  L -

          CACAGCCACGTCAGGACCGCGCACCGGGAACCCAGCCCAGCCCGTCCGTCGACGCGGCCA
    11341 ---------+---------+---------+---------+---------+---------+ 11400
          GTGTCGGTGCAGTCCTGGCGCGTGGCCCTTGGGTCGGGTCGGGCAGGCAGCTGCGCCGGT
    0-<      V  A  V -
    11-*                   *  S  R  A  G  P  V  W  G  L  G  D  T  S  A  A -

          GACCCCCCTGCCTCACCGGTCGCTCGGCCCCCGCCTCATCCCCCAGAAGAGCCCGTGCCT
    11401 ---------+---------+---------+---------+---------+---------+ 11460
          CTGGGGGGACGGAGTGGCCAGCGAGCCGGGGGCGGAGTAGGGGGTCTTCTCGGGCACGGA
    11       L  G  G  Q  R  V  P  R  E  A  G  A  E  D  G  L  L  A  R  A -

          GCAGTGCGGCGCTCTGCTCCATGAGGCGGCCCACCACCTTTCCCGGCACGGCGCCGTGCG
    11461 ---------+---------+---------+---------+---------+---------+ 11520
          CGTCACGCCGCGAGACGAGGTACTCCGCCGGGTGGTGGAAAGGGCCGTGCCGCGGCACGC
    11       Q  L  A  A  S  Q  E  M  L  R  G  V  V  K  G  P  V  A  G  H -

          GCCCGTCGGCGTCGCCCGCAGCGGTGTGCGTCATGCCGGCCATCTCGTCGGACGCCTCGG
    11521 ---------+---------+---------+---------+---------+---------+ 11580
          CGGGCAGCCGCAGCGGGCGTCGCCACACGCAGTACGGCCGGTAGAGCAGCCTGCGGAGCC
    11       P  G  D  A  D  G  A  A  T  H  T  M  G  A  M  E  D  S  A  E -

          AGAACCGCTGCCTGGCCCGGGCCGTGTCGGCGAACTCGTCGGAGGAGACCCCGCCGATCA
    11581 ---------+---------+---------+---------+---------+---------+ 11640
          TCTTGGCGACGGACCGGGCCCGGCACAGCCGCTTGAGCAGCCTCCTCTGGGGCGGCTAGT
    11       S  F  R  Q  R  A  R  A  T  D  A  F  E  D  S  S  V  G  G  I -

          GTTCGACGAAGGACTGCAGGTCGGAGTCCGCGGTGTTGGAGATCTTCCGGGCCTGCCAGA
    11641 ---------+---------+---------+---------+---------+---------+ 11700
          CAAGCTGCTTCCTGACGTCCAGCCTCAGGCGCCACAACCTCTAGAAGGCCCGGACGGTCT
    11       L  E  V  F  S  Q  L  D  S  D  A  T  N  S  I  K  R  A  Q  W -

          AATAGGAGTCCTCCGAATGGTGCATGTCGTAGAAGCCGACCAGGAACTCGTAGAAGCGGC
    11701 ---------+---------+---------+---------+---------+---------+ 11760
          TTATCCTCAGGAGGCTTACCACGTACAGCATCTTCGGCTGGTCCTTGAGCATCTTCGCCG
    11       F  Y  S  D  E  S  H  H  M  D  Y  F  G  V  L  F  E  Y  F  R -

          CGTACTCCAGCCGGTAGCGGGCCTCGAACTCCTCGAACGCGCTGGTCTCGTCGACCGACC
    11761 ---------+---------+---------+---------+---------+---------+ 11820
          GCATGAGGTCGGCCATCGCCCGGAGCTTGAGGAGCTTGCGCGACCAGAGCAGCTGGCTGG
    11       G  Y  E  L  R  Y  R  A  E  F  E  E  F  A  S  T  E  D  V  S -

          CGTCCAGGCAGGAGTTGAGCGAGCGCGCTGCCAGCAGTCCGCTGTAGGTGGCGAGGTGCA
    11821 ---------+---------+---------+---------+---------+---------+ 11880
          GCAGGTCCGTCCTCAACTCGCTCGCGCGACGGTCGTCAGGCGACATCCACCGCTCCACGT
    11       G  D  L  C  S  N  L  S  R  A  A  L  L  G  S  Y  T  A  L  H -

          CCCCGGAGGAGAACACCGGGTCGACGAAGCACGCGGCATCCCCGACCAGGGCCATGCCCG
    11881 ---------+---------+---------+---------+---------+---------+ 11940
          GGGGCCTCCTCTTGTGGCCCAGCTGCTTCGTGCGCCGTAGGGGCTGGTCCCGGTACGGGC
    11       V  G  S  S  F  V  P  D  V  F  C  A  A  D  G  V  L  A  M  G -

          GCGCCCAGAACTTCGTGTTGCTGTACGACCAGTCCTTGCGGACCCGGAGCTCGCCGTAGG
    11941 ---------+---------+---------+---------+---------+---------+ 12000
          CGCGGGTCTTGAAGCACAACGACATGCTGGTCAGGAACGCCTGGGCCTCGAGCGGCATCC
    11       P  A  W  F  K  T  N  S  Y  S  W  D  K  R  V  R  L  E  G  Y -

          GGCCCTCGGTCACCCGGGTGGCCTCGGAGAGCTTCTCCGCGATCAGCGGGCAGGCCGCGA
    12001 ---------+---------+---------+---------+---------+---------+ 12060
          CCGGGAGCCAGTGGGCCCACCGGAGCCTCTCGAAGAGGCGCTAGTCGCCCGTCCGGCGCT
    11       P  G  E  T  V  R  T  A  E  S  L  K  E  A  I  L  P  C  A  A -
```

15

```
          TGAACGACTCCATCGCCTTCTCGGGGTCGCCCTGCACCAGGCTCGCCGAGTCCCGGTTCA
    12061 ---------+---------+---------+---------+---------+---------+ 12120
    11     I  F  S  E  M  A  K  E  P  D  G  Q  V  L  S  A  S  D  R  N  -

          CCACTGCGCCGACACTCGTCAGCTCGGGAGACAGGGGTATGTACCAGAACCACCGTGCT
    12121 ---------+---------+---------+---------+---------+---------+ 12180
          GGTGACGCGGCTGTGAGCAGTCGAGCCCTCTGTCCCCATACATGGTCTTGGTGGGCACGA
    11     V  V  A  G  V  S  T  L  E  P  S  L  P  I  Y  W  F  W  G  H  -

          CGAAGGTGCAGGTGAAGATGTTCCCGGAGTTCGGCTTCGGAAGCCGCTTGCCGCCGTTGA
    12181 ---------+---------+---------+---------+---------+---------+ 12240
          GCTTCCACGTCCACTTCTACAAGGGCCTCAAGCCGAAGCCTTCGGCGAACGGCGGCAACT
    11     E  F  T  C  T  F  I  N  G  S  N  P  K  P  L  R  K  G  G  N  -

          AGTAGCCGAACAGGGCCAGGTTGCGGAAGAAGGGCGAGTACTCGCGCTTGGCGCCCGACT
    12241 ---------+---------+---------+---------+---------+---------+ 12300
          TCATCGGCTTGTCCCGGTCCAACGCCTTCTTCCCGCTCATGAGCGCGAACCGCGGGCTGA
    11     F  Y  G  F  L  A  L  N  R  F  F  P  S  Y  E  R  K  A  G  S  -

          TCTTGTACAGCCCACCGGTGTTGCCGGAGGCGTCCACGACGAAACGGGAGCCCACCTCGT
    12301 ---------+---------+---------+---------+---------+---------+ 12360
          AGAACATGTCGGGTGGCCACAACGGCCTCCGCAGGTGCTGCTTTGCCCTCGGGTGGAGCA
    11     K  K  Y  L  G  G  T  N  G  S  A  D  V  V  F  R  S  G  V  E  -

          GCTCGCGCCCCTCGGAGTCCCGGTAGCGCACGCCCCGCACCCGGCCGTCCTCGGCCTTGA
    12361 ---------+---------+---------+---------+---------+---------+ 12420
          CGAGCGCGGGGAGCCTCAGGGCCATCGCGTGCGGGGCGTGGGCCGGCAGGAGCCGGAACT
    11     H  E  R  G  E  S  D  R  Y  R  V  G  R  V  R  G  D  E  A  K  -

          GCACGTCGAGGACATCGCTGTTCTCCCGCACCTCGACACCGTGCCTGCGAGCGTTGTCGA
    12421 ---------+---------+---------+---------+---------+---------+ 12480
          CGTGCAGCTCCTGTAGCGACAAGAGGGCGTGGAGCTGTGGCACGGACGCTCGCAACAGCT
    11     L  V  D  L  V  D  S  N  E  R  V  E  V  G  H  R  R  A  N  D  -

          GCAGGATCTGGTCGAACTTCATGCGCTCGACCTGGTACGCGTACCCCGTCGCCCCCGGCA
    12481 ---------+---------+---------+---------+---------+---------+ 12540
          CGTCCTAGACCAGCTTGAAGTACGCGAGCTGGACCATGCGCATGGGGCAGCGGGGGCCGT
    11     L  L  I  Q  D  F  K  M  R  E  V  Q  Y  A  Y  G  T  A  G  P  -

          TCCGGCGCGAGACGGCGAAGTCGAACGTCCACGGTTCGGGGTTGGCACCCCACTTGAACG
    12541 ---------+---------+---------+---------+---------+---------+ 12600
          AGGCCGCGCTCTGCCGCTTCAGCTTGCAGGTGCCAAGCCCCAACCGTGGGGTGAACTTGC
    11     M  R  R  S  V  A  F  D  F  T  W  P  E  P  N  A  G  W  K  F  -

          TCCCGCCGTGCTTGATCGTGAAGGCTGCCTTCTTCAGCTCGTCGGAGACACCGAGGAGGT
    12601 ---------+---------+---------+---------+---------+---------+ 12660
          AGGGCGGCACGAACTAGCACTTCCGACGGAAGAAGTCGAGCAGCCTCTGTGGCTCCTCCA
    11     T  G  G  H  K  I  T  F  A  A  K  K  L  E  D  S  V  G  L  L  -

          GTGCGATGCCGTGGACGGTGGAGGGGAGGAGCGACTCACCGATCTGGTAGCGCGGGAAGG
    12661 ---------+---------+---------+---------+---------+---------+ 12720
          CACGCTACGGCACCTGCCACCTCCCCTCCTCGCTGAGTGGCTAGACCATCGCGCCCTTCC
    11     H  A  I  G  H  V  T  S  P  L  L  S  E  G  I  Q  Y  R  P  F  -

          TCTCCTTCTCCAGCTGGAGTACGCGATGGCCCCGCTTGCGGACCAGCGTGGAGACGGTCG
    12721 ---------+---------+---------+---------+---------+---------+ 12780
          AGAGGAAGAGGTCGACCTCATGCGCTACCGGGGCGAACGCCTGGTCGCACCTCTGCCAGC
    11     T  E  K  E  L  Q  L  V  R  H  G  R  K  R  V  L  T  S  V  T  -

          AGCCCGCCGGACCTCCGCCGACCACGATGACGTCGTACTGCGCTGACACGTCCACGGACT
    12781 ---------+---------+---------+---------+---------+---------+ 12840
          TCGGGCGGCCTGGAGGCGGCTGGTGCTACTGCAGCATGACGCGACTGTGCAGGTGCCTGA
    11-<    S  G  A  P  G  G  G  V  V  I  V  D  Y  Q  A  S  V  D  M  -

          CTCCTTCTCGCACATCGGGCGTCTCATATTCCCAGGAATCCTCTGGCCCGCCCAGGTGCT
```

```
                    ●                              ●

12841  ---------+---------+---------+---------+---------+---------+  12900
       GAGGAAGAGCGTGTAGCCCGCAGAGTATAAGGGTCCTTAGGAGACCGGGCGGGTCCACGA


       GCCGCATCTTCGGTATTGCGAAGTCGTGGGCATTCTGCGAGAAGCATGAACCGCGTGGCC
12901  ---------+---------+---------+---------+---------+---------+  12960
       CGGCGTAGAAGCCATAACGCTTCAGCACCCGTAAGACGCTCTTCGTACTTGGCGCACCGG


       CGGTCTACAGTGGCGTGGAATTTCAGTGATTGCGCTGAAGGGCGGCACACGATGAAGGCA
12961  ---------+---------+---------+---------+---------+---------+  13020
       GCCAGATGTCACCGCACCTTAAAGTCACTAACGCGACTTCCCGCCGTGTGCTACTTCCGT
10->                                                        M  K  A   -


       CTTGTACTGTCGGGTGGTTCGGGGACCCGCCTGCGCCCGATCAGTTACGCCATGCCGAAG
13021  ---------+---------+---------+---------+---------+---------+  13080
       GAACATGACAGCCCACCAAGCCCCTGGGCGGACGCGGGCTAGTCAATGCGGTACGGCTTC
10      L  V  L  S  G  G  S  G  T  R  L  R  P  I  S  Y  A  M  P  K   -


       CAGCTCGTTCCGATCGCCGGGAAGCCAGTCCTTGAATATGTTCTGGATAATATCCGGAAC
13081  ---------+---------+---------+---------+---------+---------+  13140
       GTCGAGCAAGGCTAGCGGCCCTTCGGTCAGGAACTTATACAAGACCTATTATAGGCCTTG
10      Q  L  V  P  I  A  G  K  P  V  L  E  Y  V  L  D  N  I  R  N   -


       CTCGATATCAAAGAGGTCGCCATTGTCGTCGGTGACTGGGCTCAGGAAATTATTGAGGCA
13141  ---------+---------+---------+---------+---------+---------+  13200
       GAGCTATAGTTTCTCCAGCGGTAACAGCAGCCACTGACCCGAGTCCTTTAATAACTCCGT
10      L  D  I  K  E  V  A  I  V  V  G  D  W  A  Q  E  I  I  E  A   -


       ATGGGTGACGGCAGCCGTTTCGGTCTGCGCCTCACCTACATACGCCAGGAGCAACCTCTG
13201  ---------+---------+---------+---------+---------+---------+  13260
       TACCCACTGCCGTCGGCAAAGCCAGACGCGGAGTGGATGTATGCGGTCCTCGTTGGAGAC
10      M  G  D  G  S  R  F  G  L  R  L  T  Y  I  R  Q  E  Q  P  L   -


       GGCATCGCGCACTGCGTGAAACTGGCCCGAGACTTCCTCGACGAGGACGACTTCGTCCTC
13261  ---------+---------+---------+---------+---------+---------+  13320
       CCGTAGCGCGTGACGCACTTTGACCGGGCTCTGAAGGAGCTGCTCCTGCTGAAGCAGGAG
10      G  I  A  H  C  V  K  L  A  R  D  F  L  D  E  D  D  F  V  L   -


       TACCTAGGCGACATCATGCTGGACGGAGACCTGTCCGCGCAGGCGGGGCACTTCCTCCAC
13321  ---------+---------+---------+---------+---------+---------+  13380
       ATGGATCCGCTGTAGTACGACCTGCCTCTGGACAGGCGCGTCCGCCCCGTGAAGGAGGTG
10      Y  L  G  D  I  M  L  D  G  D  L  S  A  Q  A  G  H  F  L  H   -


       ACCCGCCCCGCCGCGCGGATCGTCGTGCGCCAGGTGCCCGACCCCCGGGCCTTCGGGGTG
13381  ---------+---------+---------+---------+---------+---------+  13440
       TGGGCGGGGCGGCGCGCCTAGCAGCACGCGGTCCACGGGCTGGGGGCCCGGAAGCCCCAC
10      T  R  P  A  A  R  I  V  V  R  Q  V  P  D  P  R  A  F  G  V   -


       ATCGAGCTGGACGGCGAAGGGCGTGTGCTGCGCCTGGTCGAGAAACCCCGTGAACCGCGC
13441  ---------+---------+---------+---------+---------+---------+  13500
       TAGCTCGACCTGCCGCTTCCCGCACACGACGCGGACCAGCTCTTTGGGGCACTTGGCGCG
10      I  E  L  D  G  E  G  R  V  L  R  L  V  E  K  P  R  E  P  R   -


       AGCGACCTCGCGGCGGTCGGCGTGTACTTCTTCACCGCGGACGTGCACCGCGCCGTCGAC
13501  ---------+---------+---------+---------+---------+---------+  13560
       TCGCTGGAGCGCCGCCAGCCGCACATGAAGAAGTGGCGCCTGCACGTGGCGCGGCAGCTG
10      S  D  L  A  A  V  G  V  Y  F  F  T  A  D  V  H  R  A  V  D   -


       GCGATTAGCCCGAGCCGACGGGGCGAGCTGGAAATCACCGACGCCATCCAGTGGCTGCTG
13561  ---------+---------+---------+---------+---------+---------+  13620
       CGCTAATCGGGCTCGGCTGCCCCGCTCGACCTTTAGTGGCTGCGGTAGGTCACCGACGAC
10      A  I  S  P  S  R  R  G  E  L  E  I  T  D  A  I  Q  W  L  L   -


       GAGCAGGGCCTGCCGGTCGAGGCCGGCCGCTACACGGACTACTGGAAGGACACCGGCCGG
13621  ---------+---------+---------+---------+---------+---------+  13680
       CTCGTCCCGGACGGCCAGCTCCGGCCGGCGATGTGCCTGATGACCTTCCTGTGGCCGGCC
10      E  Q  G  L  P  V  E  A  G  R  Y  T  D  Y  W  K  D  T  G  R
```

17

```
              GTCGAGGACGTCGTGGAGTGCAACCGGCGGATGCTCGGCCGTCTGGCGCTCCAGGTGTCG
    13681     ---------+---------+---------+---------+---------+---------+ 13740
              CAGCTCCTGCAGCACCTCACGTTGGCCGCCTACGAGCCGGCAGACCGCGAGGTCCACAGC
    10         V   E   D   V   V   E   C   N   R   R   M   L   G   R   L   A   L   Q   V   S   -

              GGCGAGGTGGACCCGGAGAGCGAACTGGTGGGTGCGGTGGTCGTCGAGGAGGGCGCCCGG
    13741     ---------+---------+---------+---------+---------+---------+ 13800
              CCGCTCCACCTGGGCCTCTCGCTTGACCACCCACGCCACCAGCAGCTCCTCCCGCGGGCC
    10         G   E   V   D   P   E   S   E   L   V   G   A   V   V   V   E   E   G   A   R   -

              GTGACGCGTTCGCGGGTCGTGGGACCAGCGGTGATCGGCGCGGGCACGGTCGTCGAGGAC
    13801     ---------+---------+---------+---------+---------+---------+ 13860
              CACTGCGCAAGCGCCCAGCACCCTGGTCGCCACTAGCCGCGCCCGTGCCAGCAGCTCCTG
    10         V   T   R   S   R   V   V   G   P   A   V   I   G   A   G   T   V   V   E   D   -

              AGCCAGATCGGACCGTACGCCTCCATCGGCCGGCGCTGCACCGTGCGGGCGTCCCGGCTC
    13861     ---------+---------+---------+---------+---------+---------+ 13920
              TCGGTCTAGCCTGGCATGCGGAGGTAGCCGGCCGCGACGTGGCACGCCCGCAGGGCCGAG
    10         S   Q   I   G   P   Y   A   S   I   G   R   R   C   T   V   R   A   S   R   L   -

              TCCGACTCCATCGTCCTTGACGACGCCTCGATCCTCGCGGTGAGCGGACTGCACGGCTCG
    13921     ---------+---------+---------+---------+---------+---------+ 13980
              AGGCTGAGGTAGCAGGAACTGCTGCGGAGCTAGGAGCGCCACTCGCCTGACGTGCCGAGC
    10         S   D   S   I   V   L   D   D   A   S   I   L   A   V   S   G   L   H   G   S   -

              CTGATCGGAAGGGGCGCGCGGATCGCGCCCGGGGGCCCGGGGCGAGGCCCGGCACCGGCTG
    13981     ---------+---------+---------+---------+---------+---------+ 14040
              GACTAGCCTTCCCCGCGCGCCTAGCGCGGGCCCCGGGCCCCGCTCCGGGCCGTGGCCGAC
    10         L   I   G   R   G   A   R   I   A   P   G   A   R   G   E   A   R   H   R   L   -

              GTCGTCGGCGACCACGTGCAGATCGAGATCGCGGCCTGACGCACCCACCGGAGCACCGGG
    14041     ---------+---------+---------+---------+---------+---------+ 14100
              CAGCAGCCGCTGGTGCACGTCTAGCTCTAGCGCCGGACTGCGTGGGTGGCCTCGTGGCCC
    10-*       V   V   G   D   H   V   Q   I   E   I   A   A   *   -

              GGGAGGCTCGGCAGGGGCGTCAGGCCGTAAGAAGGGCTGCCGGGGCGGGACGGACCCGCC
    14101     ---------+---------+---------+---------+---------+---------+ 14160
              CCCTCCGAGCCGTCCCCGCAGTCCGGCATTCTTCCCGACGGCCCCGCCCTGCCTGGGCGG

              CCGGCAGCCCACAGGTCCCCGGTCCGCGGATATGGGGGACTCGAGGTTCGATCAGCCGAA
    14161     ---------+---------+---------+---------+---------+---------+ 14220
              GGCCGTCGGGTGTCCAGGGGCCAGGCGCCTATACCCCCTGAGCTCCAAGCTAGTCGGCTT
    9-*                                                             *   G   F   -

              GGTCAGAGCCACGTGGCCGAGGTCGAGCCCGGAGTTGCCGGCGCCGAGGTTACAGGCGGC
    14221     ---------+---------+---------+---------+---------+---------+ 14280
              CCAGTCTCGGTGCACCGGCTCCAGCTCGGGCCTCAACGGCCGCGGCTCCAATGTCCGCCG
    9              T   L   A   V   H   G   L   D   L   G   S   N   G   A   G   L   N   C   A   A   -

              CGTGGCGCAGTCGACGCTGCCGACCGGCGTGCCTTCGGGCGTGGAGCCCGTGTACGACTT
    14281     ---------+---------+---------+---------+---------+---------+ 14340
              GCACCGCGTCAGCTGCGACGGCTGGCCGCACGGAAGCCCGCACCTCGGGCACATGCTGAA
    9              T   A   C   D   V   S   G   V   P   T   G   E   P   T   S   G   T   Y   S   K   -

              GCGCACGACGAAGCTGAACGACGCCGCTCCGGACGCGTCCGTGGTGAAGGACGTCGCGGT
    14341     ---------+---------+---------+---------+---------+---------+ 14400
              CGCGTGCTGCTTCGACTTGCTGCGGCGAGGCCTGCGCAGGCACCACTTCCTGCAGCGCCA
    9              R   V   V   F   S   F   S   A   A   G   S   A   D   T   T   F   S   T   A   T   -

              CGCCGGGTTGCACGCGTCCTGGCCACCGACCGGAGCGCACTGGGCGATGTAGTAGGTCTC
    14401     ---------+---------+---------+---------+---------+---------+ 14460
              GCGGCCCAACGTGCGCAGGACCGGTGGCTGGCCTCGCGTGACCCGCTACATCATCCAGAG
    9              A   P   N   C   A   D   Q   G   G   V   P   A   C   Q   A   I   Y   Y   T   E   -

              GCCGGCGGCGGCACCGCTGACCGACACCGACACGCTCTGTCCGTCACTCAGACCCGAGGC
```

18

```
                                      ●                         ●


             14461 ---------+---------+---------+---------+---------+---------+ 14520
                   CGGCCGCCGCCGTGGCGACTGGCTGTGGCTGTGCGAGACAGGCAGTGAGTCTGGGCTCCG
             9       G  A  A  A  G  S  V  S  V  S  V  S  Q  G  D  S  L  G  S  A  -


                   GGGACTGACGGAGAAGGCGGGCGCGGCGAAGGCGACGGACTGTGCGGCGGCGGCCAGGCC
             14521 ---------+---------+---------+---------+---------+---------+ 14580
                   CCCTGACTGCCTCTTCCGCCCGCGCCGCTTCCGCTGCCTGACACGCCGCCGCCGGTCCGG
             9       P  S  V  S  F  A  P  A  A  F  A  V  S  Q  A  A  A  A  L  G  -


                   GATGGATGCGACGGCCACGACGCCGAACCTGGAAGCACGGCGGGACATGTGACGTAACGA
             14581 ---------+---------+---------+---------+---------+---------+ 14640
                   CTACCTACGCTGCCGGTGCTGCGGCTTGGACCTTCGTGCCGCCCTGTACACTGCATTGCT
             9       I  S  A  V  A  V  V  G  F  R  S  A  R  R  S  M  H  R  L  S  -


                   CATGCGTAGGCTCCGATTCGAGGAGGGGGTTGATCACTCCATGAAAGGATCACCTCGCCG
             14641 ---------+---------+---------+---------+---------+---------+ 14700
                   GTACGCATCCGAGGCTAAGCTCCTCCCCCAACTAGTGAGGTACTTTCCTAGTGGAGCGGC
             9-<      M  -
             8-*                                                  *  R  A  -


                   GACGGCCGCCTGCATCTCCCTCTGTGCTCTCGTGGATTTCCGGCACGGCACTCCCGTCGA
             14701 ---------+---------+---------+---------+---------+---------+ 14760
                   CTGCCGGCGGACGTAGAGGGAGACACGAGAGCACCTAAAGGCCGTGCCGTGAGGGCAGCT
             8       P  R  G  G  A  D  G  E  T  S  E  H  I  E  P  V  A  S  G  D  -


                   CGGCCGCCCGCAGAATGCGGCAGACCCCCCGCACCTCCTCCGGCCCCACCGCCGTACCGG
             14761 ---------+---------+---------+---------+---------+---------+ 14820
                   GCCGGCGGGCGTCTTACGCCGTCTGGGGGGCGTGGAGGAGGCCGGGGTGGCGGCATGGCC
             8       V  A  A  R  L  I  R  C  V  G  R  V  E  E  P  G  V  A  T  G  -


                   TGGGCAGCGACAGCACCCGCTCGGTGAGCGCCTCCACCTTCGGGAGCGGATCGGGCGCGT
             14821 ---------+---------+---------+---------+---------+---------+ 14880
                   ACCCGTCGCTGTCGTGGGCGAGCCACTCGCGGAGGTGGAAGCCCTCGCCTAGCCCGCGCA
             8       T  P  L  S  L  V  R  E  T  L  A  E  V  K  P  L  P  D  P  A  -


                   GGCGCGCGAGGTCGGACCGGTAGGGCTCGCAGCTGTGGCAGCCGGGGCTGAAGTAGGCGC
             14881 ---------+---------+---------+---------+---------+---------+ 14940
                   CCGCGCGCTCCAGCCTGGCCATCCCGAGCGTCGACACCGTCGGCCCCGACTTCATCCGCG
             8       H  R  A  L  D  S  R  Y  P  E  C  S  H  C  G  P  S  F  Y  A  -


                   GGGCCAGGACGTTGTGCCGTTGGAGCACCGCCTGGAGTTCGTCGCGGTGCAGCCCGGCGC
             14941 ---------+---------+---------+---------+---------+---------+ 15000
                   CCCGGTCCTGCAACACGGCAACCTCGTGGCGGACCTCAAGCAGCGCCACGTCGGGCCGCG
             8       R  A  L  V  N  H  R  Q  L  V  A  Q  L  E  D  R  H  L  G  A  -


                   GGACGGCGTCCACCTCGATGACGACGTACTGGCAGTTCGACAGCTCGTTCGGATCCTGCG
             15001 ---------+---------+---------+---------+---------+---------+ 15060
                   CCTGCCGCAGGTGGAGCTACTGCTGCATGACCGTCAAGCTGTCGAGCAAGCCTAGGACGC
             8       R  V  A  D  V  E  I  V  V  Y  Q  C  N  S  L  E  N  P  D  Q  -


                   GGCGGACCCGGACGCCGGGCAGTCCGTCGAGGTACTGCTCGTACAGACGGTAGTTGCGCC
             15061 ---------+---------+---------+---------+---------+---------+ 15120
                   CCGCCTGGGCCTGCGGCCCGTCAGGCAGCTCCATGACGAGCATGTCTGCCATCAACGCGG
             8       P  R  V  R  V  G  P  L  G  D  L  Y  Q  E  Y  L  R  Y  N  R  -


                   GGTTGATCGCGGTGAAGTGATCGGCGGACTCCAGGGAGGTGAGGCCCATGGCCGCGCTGA
             15121 ---------+---------+---------+---------+---------+---------+ 15180
                   CCAACTAGCGCCACTTCACTAGCCGCCTGAGGTCCCTCCACTCCGGGTACCGGCGCGACT
             8       R  N  I  A  T  F  H  D  A  S  E  L  S  T  L  G  M  A  A  S  -


                   TCTCGTGCATCCGCGCGACCGTTCCGCTCCCGGTGATCTCATGCGCGGCGTTGAGCCCCT
             15181 ---------+---------+---------+---------+---------+---------+ 15240
                   AGAGCACGTAGGCGCGCTGGCAAGGCGAGGGCCACTAGAGTACGCGCCGCAACTCGGGGA
             8       I  E  H  M  R  A  V  T  G  S  G  T  I  E  H  A  A  N  L  G  -


                   GGTGGCGCATGGCCCGGAGCCGGTCGGCCAGGGCGTCGTCGTCGGTGACGATCGCCCCGC
```

19

```
          15241 ---------+---------+---------+---------+---------+---------+ 15300
                CCACCGCGTACCGGGCCTCGGCCAGCCGGTCCCGCAGCAGCAGCCACTGCTAGCGGGGCG
       8          Q  H  R  M  A  R  L  R  D  A  L  A  D  D  D  T  V  I  A  G  -

                CCTCGAAGCTGTTCACGAACTTCGTCGCCTGGAAGCTGAAGATCTCCGCCGTGCCGAAGC
          15301 ---------+---------+---------+---------+---------+---------+ 15360
                GGAGCTTCGACAAGTGCTTGAAGCAGCGGACCTTCGACTTCTAGAGGCGGCACGGCTTCG
       8          G  E  F  S  N  V  F  K  T  A  Q  F  S  F  I  E  A  T  G  F  -

                CGCCGATCGGCTTCGACCGGTAGGTGCAGCCGAAGGCGTGGGCGGCATCGAAGAGCAGGT
          15361 ---------+---------+---------+---------+---------+---------+ 15420
                GCGGCTAGCCGAAGCTGGCCATCCACGTCGGCTTCCGCACCCGCCGTAGCTTCTCGTCCA
       8          G  G  I  P  K  S  R  Y  T  C  G  F  A  H  A  A  D  F  L  L  -

                GCAGCCCGTGCTCGGCGGCCAGCTTGGTCAGCTCGTCGATCCGGGCCGGTCTGCCGAAGA
          15421 ---------+---------+---------+---------+---------+---------+ 15480
                CGTCGGGCACGAGCCGCCGGTCGAACCAGTCGAGCAGCTAGGCCCGGCCAGACGGCTTCT
       8          H  L  G  H  E  A  A  L  K  T  L  E  D  I  R  A  P  R  G  F  -

                CGTGCACGTCCAGGATGGCGCGGGTACGCGGGCCGATGAGCCGCTCCACGTGTGCCACGT
          15481 ---------+---------+---------+---------+---------+---------+ 15540
                GCACGTGCAGGTCCTACCGCGCCCATGCGCCCGGCTACTCGGCGAGGTGCACACGGTGCA
       8          V  H  V  D  L  I  A  R  T  R  P  G  I  L  R  E  V  H  A  V  -

                CCGCGGTTCCGGTCTCCTCGTCCAGTTCGCAGAAGACAGGCACCGCACCGATCCAGTCCA
          15541 ---------+---------+---------+---------+---------+---------+ 15600
                GGCGCCAAGGCCAGAGGAGCAGGTCAAGCGTCTTCTGTCCGTGGCGTGGCTAGGTCAGGT
       8          D  A  T  G  T  E  E  D  L  E  C  F  V  P  V  A  G  I  W  D  -

                GTGCGTGGGCGGTGGCGACCCAGGTGAAGGAGGGCACGATCACCTCGTCCCCAGGACCGA
          15601 ---------+---------+---------+---------+---------+---------+ 15660
                CACGCACCCGCCACCGCTGGGTCCACTTCCTCCCGTGCTAGTGGAGCAGGGGTCCTGGCT
       8          L  A  H  A  T  A  V  W  T  F  S  P  V  I  V  E  D  G  P  G  -

                TGCCCAGGGCCTTCGCGGCGACCTGGATGCCGGTGGTGGCGTTCGATACGGCGACGCAGT
          15661 ---------+---------+---------+---------+---------+---------+ 15720
                ACGGGTCCCGGAAGCGCCGCTGGACCTACGGCCACCACCGCAAGCTATGCCGCTGCGTCA
       8          I  G  L  A  K  A  A  V  Q  I  G  T  T  A  N  S  V  A  V  C  -

                GCCTGACCTGGGTCAGCTCGGCCACACGGGCCTCGAACTCCCGGACCAGGGGGCCGTCAT
          15721 ---------+---------+---------+---------+---------+---------+ 15780
                CGGACTGGACCCAGTCGAGCCGGTGTGCCCGGAGCTTGAGGGCCTGGTCCCCCGGCAGTA
       8          H  R  V  Q  T  L  E  A  V  R  A  E  F  E  R  V  L  P  G  D  -

                TGGTGAACCACAGGCGCTCCAGCGCCCCGTCGATCCGTTCCATCAAACGGTCGCGGGAGC
          15781 ---------+---------+---------+---------+---------+---------+ 15840
                ACCACTTGGTGTCCGCGAGGTCGCGGGGCAGCTAGGCAAGGTAGTTTGCCAGCGCCCTCG
       8          N  T  F  W  L  R  E  L  A  G  D  I  R  E  M  L  R  D  R  S  -

                CCACGTTCGGGCGTCCCACGTGCAGCGGTTCGCTGAAGTAGGGCGTGGGTAGGGAGTCCA
          15841 ---------+---------+---------+---------+---------+---------+ 15900
                GGTGCAAGCCCGCAGGGTGCACGTCGCCAAGCGACTTCATCCCGCACCCATCCCTCAGGT
       8          G  V  N  P  R  G  V  H  L  P  E  S  F  Y  P  T  P  L  S  D  -

                GACGCACCGGGCCGCCGCTCATGCCGTGCGCACGCCGACGAAGAGGCCGGGGCTGTTGGG
          15901 ---------+---------+---------+---------+---------+---------+ 15960
                CTGCGTGGCCCGGCGGCGAGTACGGCACGCGTGCGGCTGCTTCTCCGGCCCCGACAACCC
       a          D  A  P  G  R  R  S  C  R  A  H  A  D  E  E  A  G  A  V  G  -
       b           T  H  R  A  A  A  H  A  V  R  T  P  T  K  R  P  G  L  L  G -
       c            R  T  G  P  P  L  M  P  C  A  R  R  R  R  G  R  G  C  W  A -
          15901 ---------+---------+---------+---------+---------+---------+ 15960
       7-*                           *  A  T  R  V  G  V  F  L  G  P  S  N  P -
       8-<         L  R  V  P  G  G  S  M

                CCGGCCGTCGGCCAGCCGGAAGCCGGGCACGAACCGCACCGAGAGCCCCACCGATTCGAA
          15961 ---------+---------+---------+---------+---------+---------+ 16020
```

```
                    GGCCGGCAGCCGGTCGGCCTTCGGCCCGTGCTTGGCGTGGCTCTCGGGGTGGCTAAGCTT
     7               R  G  D  A  L  R  F  G  P  V  F  R  V  S  L  G  V  S  E  F -

                    GGCGTCGGTGTACTGCTCGCGGGTGAAGAGGCTGGAGGTCAGGACCTCGGAGAACTCTCT
     16021      ---------+---------+---------+---------+---------+---------+   16080
                    CCGCAGCCACATGACGAGCGCCCACTTCTCCGACCTCCAGTCCTGGAGCCTCTTGAGAGA
     7               A  D  T  Y  Q  E  R  T  F  L  S  S  T  L  V  E  S  F  E  R -

                    GAAGCCGGAGGCGTCCGCGACCCGGAACCGGACCTCCAGACGTGACTTGTCGCCCTGGCG
     16081      ---------+---------+---------+---------+---------+---------+   16140
                    CTTCGGCCTCCGCAGGCGCTGGGCCTTGGCCTGGAGGTCTGCACTGAACAGCGGGACCGC
     7               F  G  S  A  D  A  V  R  F  R  V  E  L  R  S  K  D  G  Q  R -

                    CACGGAGTGCGTCATCCGCGTGATGACACGGCCCTCCTCCTGGTGCAGATGGCCGCCGAC
     16141      ---------+---------+---------+---------+---------+---------+   16200
                    GTGCCTCACGCAGTAGGCGCACTACTGTGCCGGGAGGAGGACCACGTCTACCGGCGGCTG
     7               V  S  H  T  M  R  T  I  V  R  G  E  E  Q  H  L  H  G  G  V -

                    ATGCCCGTCGAGGAAGTTCTCGGGGAAATACCAGGGGTTCGGCGACGAGGACTCCCCCGGG
     16201      ---------+---------+---------+---------+---------+---------+   16260
                    TACGGGCAGCTCCTTCAAGAGCCCCTTTATGGTCCCAAGCCGCTGCTCCTGAGGGGGCCC
     7               H  G  D  L  F  N  E  P  F  Y  W  P  E  A  V  L  V  G  G  P -

                    GTTCAGGTGGTGGGCCATGGCCGACACCGCGGCCTTGAGCTCGGTGACGGACCCCATCTC
     16261      ---------+---------+---------+---------+---------+---------+   16320
                    CAAGTCCACCACCCGGTACCGGCTGTGGCGCCGGAACTCGAGCCACTGCCTGGGGTAGAG
     7               N  L  H  H  A  M  A  S  V  A  A  K  L  E  T  V  S  G  M  E -

                    GCCGAGCGCGTTGCCCATGCAGGTGATCGCGTCGAAGGTGCGGCCCAGGTCGAACGAACG
     16321      ---------+---------+---------+---------+---------+---------+   16380
                    CGGCTCGCGCAACGGGTACGTCCACTAGCGCAGCTTCCACGCCGGGTCCAGCTTGCTTGC
     7               G  L  A  N  G  M  C  T  I  A  D  F  T  R  G  L  D  F  S  R -

                    CATGTCACCGGCGTGCAGCGGGACGCCGGGAAGCCGGCCCGCCGCCTGCTCCAGCATCGC
     16381      ---------+---------+---------+---------+---------+---------+   16440
                    GTACAGTGGCCGCACGTCGCCCTGCGGCCCTTCGGCCGGGCGGCGGACGAGGTCGTAGCG
     7               M  D  G  A  H  L  P  V  G  P  L  R  G  A  A  Q  E  L  M  A -

                    GGGCGCGTACTCGAGGCCCTCCACATGGCCGAAGAGCGTGGCGAGCGTCTCCAGATGGGC
     16441      ---------+---------+---------+---------+---------+---------+   16500
                    CCCGCGCATGAGCTCCGGGAGGTGTACCGGCTTCTCGCACCGCTCGCAGAGGTCTACCCG
     7               P  A  Y  E  L  G  E  V  H  G  F  L  T  A  L  T  E  L  H  A -

                    TCCGGTGCCGCAGGCGACGTCCAGGAGCGACACGGCGTCGGGGCGGGCGGCGAGGATCAG
     16501      ---------+---------+---------+---------+---------+---------+   16560
                    AGGCCACGGCGTCCGCTGCAGGTCCTCGCTGTGCCGCAGCCCCGCCCGCCGCTCCTAGTC
     7               G  T  G  C  A  V  D  L  L  S  V  A  D  P  R  A  A  L  I  L -

                    CTCGGTGAGCCCGCGGGCCTCCAGGTCGAAGTCCTTGCCGCGGCTGCGGAACACGAGGTC
     16561      ---------+---------+---------+---------+---------+---------+   16620
                    GAGCCACTCGGGCGCCCGGAGGTCCAGCTTCAGGAACGGCGCCGACGCCTTGTGCTCCAG
     7               E  T  L  G  R  A  E  L  D  F  D  K  G  R  S  R  F  V  L  D -

                    GTAGAACTTCGCGTGCTCGGGGCCGTACTCCATCAGACGAGCTCCTTCGCAGACTGGGCG
     16621      ---------+---------+---------+---------+---------+---------+   16680
                    CATCTTGAAGCGCACGAGCCCCGGCATGAGGTAGTCTGCTCGAGGAAGCGTCTGACCCGC
     7-<             Y  F  K  A  H  E  P  G  Y  E  M -
     6-*                                            *  V  L  E  K  A  S  Q  A -

                    GAGATGATTCTGGGCTCCGGGATGGGAACGATGAACTTCCCTCCCGCCTCCAGGAAGCGG
     16681      ---------+---------+---------+---------+---------+---------+   16740
                    CTCTACTAAGACCCGAGGCCCTACCCTTGCTACTTGAAGGGAGGGCGGAGGTCCTTCGCC
     6               S  I  I  R  P  E  P  I  P  V  I  F  K  G  G  A  E  L  F  R -

                    CGCTCCTTGCGGACGACCTCGTCGGTGTAGTTCCAGGCGAGGAGGAGGTAGTAGTCCGGC
     16741      ---------+---------+---------+---------+---------+---------+   16800
```

21

```
        GCGAGGAACGCCTGCTGGAGCAGCCACATCAAGGTCCGCTCCTCCTCCATCATCAGGCCG
6        R  E  K  R  V  V  E  D  T  Y  N  W  A  L  L  L  Y  Y  D  P  -

         TCGGTGGCAGCGACCTCCTCCGGAGGAAGGACCGGGATGCGGTTCCCCGGCAGCAGTTTG
16801    ---------+---------+---------+---------+---------+---------+  16860
         AGCCACCGTCGCTGGAGGAGGCCTCCTTCCTGGCCCTACGCCAAGGGGCCGTCGTCAAAC
6        E  T  A  A  V  E  E  P  P  L  V  P  I  R  N  G  P  L  L  K  -

         CCGTGCTTGAGGCTGGTGGTGTCGCCGCAGACGGTGATGTCCTGATCCGTCAGACCGCAG
16861    ---------+---------+---------+---------+---------+---------+  16920
         GGCACGAACTCCGACCACCACAGCGGCGTCTGCCACTACAGGACTAGGCAGTCTGGCGTC
6        G  H  K  L  S  T  T  D  G  C  V  T  I  D  Q  D  T  L  G  C  -

         GCCATCAGCAACTGGGTCCCCTTGGACGGTGCTCCGTAGCCGGCCACGCGGTGGCCGTCC
16921    ---------+---------+---------+---------+---------+---------+  16980
         CGGTAGTCGTTGACCCAGGGGAACCTGCCACGAGGCATCGGCCGGTGCGCCACCGGCAGG
6        A  M  L  L  Q  T  G  K  S  P  A  G  Y  G  A  V  R  H  G  D  -

         GCGGCCAGACCGCGAACGAGCGTACGGATCGCTTCGGTCACGCGCGTCACCCGCTCGGCG
16981    ---------+---------+---------+---------+---------+---------+  17040
         CGCCGGTCTGGCGCTTGCTCGCATGCCTAGCGAAGCCAGTGCGCGCAGTGGGCGAGCCGC
6        A  A  L  G  R  V  L  T  R  I  A  E  T  V  R  T  V  R  E  A  -

         AACGCCCGGTAGGGGGCATCCGTCAGCAGTCCGCGCTCCTCCTCCAGGCCGAGCAGCGCC
17041    ---------+---------+---------+---------+---------+---------+  17100
         TTGCGGGCCATCCCCCGTAGGCAGTCGTCAGGCGCGAGGAGGAGGTCCGGCTCGTCGCGG
6        F  A  R  Y  P  A  D  T  L  L  G  R  E  E  E  L  G  L  L  A  -

         GCGACCGAGGGCTCCGGGACCCGTGCGGCCGACTCGCGCGCGGCGACGACCGCGATCGAA
17101    ---------+---------+---------+---------+---------+---------+  17160
         CGCTGGCTCCCGAGGCCCTGGGCACGCCGGCTGAGCGCGCGCCGCTGCTGGCGCTAGCTT
6        A  V  S  P  E  P  V  R  A  A  S  E  R  A  A  V  V  A  I  S  -

         CCGCCGTGCACGGCGACCCGCTCCACGTCGATGATCCGCAGGCCGTGCGCGCCGAAGAGG
17161    ---------+---------+---------+---------+---------+---------+  17220
         GGCGGCACGTGCCGCTGGGCGAGGTGCAGCTACTAGGCGTCCGGCACGCGCGGCTTCTCC
6        G  G  H  V  A  V  R  E  V  D  I  I  R  L  G  H  A  G  F  L  -

         TGGCGCAGTGTGTGCAGGGAGAAGTACGACAGGTGCTCGTGGTAGATCGTGTCGAACTGG
17221    ---------+---------+---------+---------+---------+---------+  17280
         ACCGCGTCACACACGTCCCTCTTCATGCTGTCCACGAGCACCATCTAGCACAGCTTGACC
6        H  R  L  T  H  L  S  F  Y  S  L  H  E  H  Y  I  T  D  F  Q  -

         TTCTCGTCGAGCAGGTTCAGCAGGTACGGCACCTCGATGACCAGGACGCCGTCGTCGTCG
17281    ---------+---------+---------+---------+---------+---------+  17340
         AAGAGCAGCTCGTCCAAGTCGTCCATGCCGTGGAGCTACTGGTCCTGCGGCAGCAGCAGC
6        N  E  D  L  L  N  L  L  Y  P  V  E  I  V  L  V  G  D  D  D  -

         AGCACTGCGTCGACGCCGTCCAGGATGCGGTGCACGTCGTCGATGTGCGCGAAGCACTGG
17341    ---------+---------+---------+---------+---------+---------+  17400
         TCGTGACGCAGCTGCGGCAGGTCCTACGCCACGTGCAGCAGCTACACGCGCTTCGTGACC
6        L  V  A  D  V  G  D  L  I  R  H  V  D  D  I  H  A  F  C  Q  -

         CGGCCGATGACGGCCTTGGCCCTGCCCTGCTCAAGGGCGATGCGGCCCGCGGGCTCCGGG
17401    ---------+---------+---------+---------+---------+---------+  17460
         GCCGGCTACTGCCGGAACCGGGACGGGACGAGTTCCCGCTACGCCGGGCGCCCGAGGCCC
6        R  G  I  V  A  K  A  R  G  Q  E  L  A  I  R  G  A  P  E  P  -

         CCGAAGAAGTCCGGGTCCGTGGGGATCCCCCGGGCGTTGGCGATCTCGGCGAGGTTGGCC
17461    ---------+---------+---------+---------+---------+---------+  17520
         GGCTTCTTCAGGCCCAGGCACCCCTAGGGGGCCCGCAACCGCTAGAGCCGCTCCAACCGG
6        G  F  F  D  P  D  T  P  I  G  R  A  N  A  I  E  A  L  N  A  -

         GCCGGGTCGACCCCGGCCACCCGCATGCCCGCCGCCCGGAACATCGCGAGCTGGGTGCCG
17521    ---------+---------+---------+---------+---------+---------+  17580
         CGGCCCAGCTGGGGCCGGTGGGCGTACGGGCGGCGGGCCTTGTAGCGCTCGACCCACGGC
```

22

```
6              A   P   D   V   G   A   V   R   M   G   A   A   R   F   M   A   L   Q   T   G   -
               ACGTTGCTGCCCAGCTCCACGACCAGGTCGCCGGAGGCGAGGCTTGCCCGGCGGGTCGCC
     17581  ---------+---------+---------+---------+---------+---------+  17640
               TGCAACGACGGGTCGAGGTGCTGGTCCAGCGGCCTCCGCTCCGAACGGGCCGCCCAGCGG
6              V   N   S   G   L   E   V   V   L   D   G   S   A   L   S   A   R   R   T   A   -

               AGCCCGACGATGTGCGCCATGTGCTCGCGGATCTGGTCGGAGTCGGAGGAGACGTAGACG
     17641  ---------+---------+---------+---------+---------+---------+  17700
               TCGGGCTGCTACACGCGGTACACGAGCGCCTAGACCAGCCTCAGCCTCCTCTGCATCTGC
6               L   G   V   I   H   A   M   H   E   R   I   Q   D   S   D   S   S   V   Y   V   -

               TAGTGCTTGAACAGTGTCCCGGGGTCGACGACATGGCGAAGCGTCATCAGCCGGCACGAC
     17701  ---------+---------+---------+---------+---------+---------+  17760
               ATCACGAACTTGTCACAGGGCCCCAGCTGCTGTACCGCTTCGCAGTAGTCGGCCGTGCTG
6               Y   H   K   F   L   T   G   P   D   V   V   H   R   L   T   M   L   R   C   S   -

               CGGCACACGATGACGTCGAGCGGGAAGACGTCCTGCGCCTCATCGGCGTCGGCCGGATCG
     17761  ---------+---------+---------+---------+---------+---------+  17820
               GCCGTGTGCTACTGCAGCTCGCCCTTCTGCAGGACGCGGAGTAGCCGCAGCCGGCCTAGC
6               R   C   V   I   V   D   L   P   F   V   D   Q   A   E   D   A   D   A   P   D   -

               ACGAACCCGTTGGCCAGCGGCAGCGAGCCGAAGGAGATCACCTCGGTCCAGTCGTCCGCA
     17821  ---------+---------+---------+---------+---------+---------+  17880
               TGCTTGGGCAACCGGTCGCCGTCGCTCGGCTTCCTCTAGTGGAGCCAGGTCAGCAGGCGT
6               V   F   G   N   A   L   P   L   S   G   F   S   I   V   E   T   W   D   D   A   -

               CCGCATACACGGCACGTCTCGTCCCGCCTGCATTTCTCCAGCATGAAGTCTCCTGACGGC
     17881  ---------+---------+---------+---------+---------+---------+  17940
               GGCGTATGTGCCGTGCAGAGCAGGGCGGACGTAAAGAGGTCGTACTTCAGAGGACTGCCG
6-<             G   C   V   R   C   T   E   D   R   R   C   K   E   L   M   -

               GAATGCCGACGCATCGGGCCCGTCGGTCCGGGGACGGTCAATCTAGGGTTCCGGCCGACG
     17941  ---------+---------+---------+---------+---------+---------+  18000
               CTTACGGCTGCGTAGCCCGGGCAGCCAGGCCCCTGCCAGTTAGATCCCAAGGCCGGCTGC

               GGCGCTCCACTTCGTATGTGCCCTACTGGTTCAGCGGAGCGGACGGGTGAACGCCCGTAC
     18001  ---------+---------+---------+---------+---------+---------+  18060
               CCGCGAGGTGAAGCATACACGGGATGACCAAGTCGCCTCGCCTGCCCACTTGCGGGCATG
17-*                                            *   R   L   P   R   T   F   A   R   V   -

               GTCCTCGATGAGGAGCTGCGGCTGCTCCATGGCCGCGAAGTGCCCGCCGCGGTCGAACTC
     18061  ---------+---------+---------+---------+---------+---------+  18120
               CAGGAGCTACTCCTCGACGCCGACGAGGTACCGGCGCTTCACGGGCGGCGCCAGCTTGAG
17              D   E   I   L   L   Q   P   Q   E   M   A   A   F   H   G   G   R   D   F   E   -

               GGTCCACCGCGTCAGGGTCGGCAGGATGCCCTCGGCGAACGACCGGATCGGCCGGGTGGC
     18121  ---------+---------+---------+---------+---------+---------+  18180
               CCAGGTGGCGCAGTCCCAGCCGTCCTACGGGAGCCGCTTGCTGGCCTAGCCGGCCCACCG
17              T   W   R   T   L   T   P   L   I   G   E   A   F   S   R   I   P   R   T   A   -

               GTCGTCCGGGAACACCGCGACGCCGACGGGGGCCGTCAGCGGCCAGGGCCCGCCCCAGGT
     18181  ---------+---------+---------+---------+---------+---------+  18240
               CAGCAGGCCCTTGTGGCGCTGCGGCTGCCCCCGGCAGTCGCCGGTCCCGGGCGGGGTCCA
17              D   D   P   F   V   A   V   G   V   P   A   T   L   P   W   P   G   G   W   T   -

               GCGGGCGAAGTCCGCCATGCCGCGAGCCGACTCGTAGTACAACTGAGCGCTGGAACCGGC
     18241  ---------+---------+---------+---------+---------+---------+  18300
               CGCCCGCTTCAGGCGGTACGGCGCTCGGCTGAGCATCATGTTGACTCGCGACCTTGGCCG
17              R   A   F   D   A   M   G   R   A   S   E   Y   Y   L   Q   A   S   S   G   A   -

               CGTCGCGGTCAGCCAGTAGATCATCACGTGGGTGAGCAGCCGGTCCCGGGAGATGGCCTC
     18301  ---------+---------+---------+---------+---------+---------+  18360
               GCAGCGCCAGTCGGTCATCTAGTAGTGCACCCACTCGTCGGCCAGGGCCCTCTACCGGAG
17              T   A   T   L   W   Y   I   M   V   H   T   L   L   R   D   R   S   I   A   E   -
```

23

```
                   CTCCACGTTCTTGCCGCCGCTCCACTCCTGGAACTTGTCGAGAATCCAGGCGAGCTGGCC
        18361      ---------+---------+---------+---------+---------+---------+ 18420
                   GAGGTGCAAGAACGGCGGCGAGGTGAGGACCTTGAACAGCTCTTAGGTCCGCTCGACCGG
        17          E  V  N  K  G  G  S  W  E  Q  F  K  D  L  I  W  A  L  Q  G -

                   GACCGGGGAGTCGGTGAGGCCGTAGGCCAGGGTCTGCGGGCGGGTGGCCTGGATGCGCTG
        18421      ---------+---------+---------+---------+---------+---------+ 18480
                   CTGGCCCCTCAGCCACTCCGGCATCCGGTCCCAGACGCCCGCCCACCGGACCTACGCGAC
        17          V  P  S  D  T  L  G  Y  A  L  T  Q  P  R  T  A  Q  I  R  Q -

                   CCAGCCGATGCCGGTGTCGGCGAACTCCCCGCTGTGCGCCAGCTTGCCCAGGTCGCTCTC
        18481      ---------+---------+---------+---------+---------+---------+ 18540
                   GGTCGGCTACGGCCACAGCCGCTTGAGGGGCGACACGCGGTCGAACGGGTCCAGCGAGAG
        17          W  G  I  G  T  D  A  F  E  G  S  H  A  L  K  G  L  D  S  E -

                   GTCCAGGCGCCCGATGGCCTCCGGGGCGTCCTGGGGCGGGAAGGTCACCAGCATGTTCAG
        18541      ---------+---------+---------+---------+---------+---------+ 18600
                   CAGGTCCGCGGGCTACCGGAGGCCCCGCAGGACCCCGCCCTTCCAGTGGTCGTACAAGTC
        17          D  L  R  G  I  A  E  P  A  D  Q  P  P  F  T  V  L  M  N  L -

                   GTGGACGCCGGCCACGTGCTCGGGGTCGGCCAGCCCCAGCTCCAGCGAGACGACCTTTCC
        18601      ---------+---------+---------+---------+---------+---------+ 18660
                   CACCTGCGGCCGGTGCACGAGCCCCAGCCGGTCGGGGTCGAGGTCGCTCTGCTGGAAAGG
        17          H  V  G  A  V  H  E  P  D  A  L  G  L  E  L  S  V  V  K  G -

                   CCAGTCGCCGCCCTGGGCGACGTAACGCTCGTAGCCGAGGCGGTTCATCAGCTCCGCCCA
        18661      ---------+---------+---------+---------+---------+---------+ 18720
                   GGTCAGCGGCGGGACCCGCTGCATTGCGAGCATCGGCTCCGCCAAGTAGTCGAGGCGGGT
        17          W  D  G  G  Q  A  V  Y  R  E  Y  G  L  R  N  M  L  E  A  W -

                   GGCGCGTGCGATCCGCCGCACGTCCCAGCCCGGCTCGGCAGTCGGGCCGGAGAAGCCGTA
        18721      ---------+---------+---------+---------+---------+---------+ 18780
                   CCGCGCACGCTAGGCGGCGTGCAGGGTCGGGCCGAGCCGTCAGCCCGGCCTCTTCGGCAT
        17          A  R  A  I  R  R  V  D  W  G  P  E  A  T  P  G  S  F  G  Y -

                   GCCCGGCATGGAGGGGACGACGACGTGGAAGGCGTCCGCCGGGTCGCCGCCGTGCGCGCG
        18781      ---------+---------+---------+---------+---------+---------+ 18840
                   CGGGCCGTACCTCCCCTGCTGCTGCACCTTCCGCAGGCGGCCCAGCGGCGGCACGCGCGC
        17          G  P  M  S  P  V  V  V  H  F  A  D  A  P  D  G  G  H  A  R -

                   CGGGTCGCTCAGCGGCCCGATGACGTCGAGGAACTCGGCGACCGAGCCCGGCCAGCCGTG
        18841      ---------+---------+---------+---------+---------+---------+ 18900
                   GCCCAGCGAGTCGCCGGGCTACTGCAGCTCCTTGAGCCGCTGGCTCGGGCCGGTCGGCAC
        17          P  D  S  L  P  G  I  V  D  L  F  E  A  V  S  G  P  W  G  H -

                   GGTGAGGATCAGCGGGATCGCGTCCGGCTCGGGCGAACGCACGTGAAGGAAGTGCACGTC
        18901      ---------+---------+---------+---------+---------+---------+ 18960
                   CCACTCCTAGTCGCCCTAGCGCAGGCCGAGCCCGCTTGCGTGCACTTCCTTCACGTGCAG
        17          T  L  I  L  P  I  A  D  P  E  P  S  R  V  H  L  F  H  V  D -

                   GGCGCCGTCGATCGTGGTGACGAACTGGGGGAACGCGTTCAGCTCGGCCTCCGCGGCACG
        18961      ---------+---------+---------+---------+---------+---------+ 19020
                   CCGCGGCAGCTAGCACCACTGCTTGACCCCCTTGCGCAAGTCGAGCCGGAGGCGCCGTGC
        17          A  G  D  I  T  T  V  F  Q  P  F  A  N  L  E  A  E  A  A  R -

                   CCAGTCGTAGCCGTGGCGCCAGTGGTCGGTGAGCTCCTTGAGGTAGGACAGCGGCACTCC
        19021      ---------+---------+---------+---------+---------+---------+ 19080
                   GGTCAGCATCGGCACCGCGGTCACCAGCCACTCGAGGAACTCCATCCTGTCGCCGTGAGG
        17          W  D  Y  G  H  R  W  H  D  T  L  E  K  L  Y  S  L  P  V  G -

                   GCGGTCCCATCCGGATCCGGGTATCTCGGACGGCCACCGGGTCGCGTCGATCCGCCGGGT
        19081      ---------+---------+---------+---------+---------+---------+ 19140
                   CGCCAGGGTAGGCCTAGGCCCATAGAGCCTGCCGGTGGCCCAGCGCAGCTAGGCGGCCCA
        17          R  D  W  G  S  G  P  I  E  S  P  W  R  T  A  D  I  R  R  T -

                   TAAGGTCGTCGAATGTCGGACTGGGTCGATCTCGATACGGAAGGGACGCACAGTGAATCC
```

24

```
                ATTCCAGCAGCTTACAGCCTGACCCAGCTAGAGCTATGCCTTCCCTGCGTGTCACTTAGG
    19141  ---------+---------+---------+---------+---------+---------+  19200
    17-<       L  T  T  S  H  R  V  P  D  I  E  I  R  F  P  R  M  -

                ACCCTCGTGATTGTGGGAGCGGGGCGGCGCGAGGCGGCCGCCCCGATGTGATCCGGGGAC
    19201  ---------+---------+---------+---------+---------+---------+  19260
                TGGGAGCACTAACACCCTCGCCCCGCCGCGCTCCGCCGGCGGGGCTACACTAGGCCCCTG

                CGTGTCTCAGGCCGGTTCGGCCGGCGCGGCCGCGCCTTCCCGTGCGGAGAAGGACCGCAC
    19261  ---------+---------+---------+---------+---------+---------+  19320
                GCACAGAGTCCGGCCAAGCCGGCCGCGCCGGCGCGGAAGGGCACGCCTCTTCCTGGCGTG
    16-*          *  A  P  E  A  P  A  A  A  G  E  R  A  S  F  S  R  V  -

                GGAGGACAGGAAGTTGCGGATCATCGGCATGCCGTGTTCGGTCCGGAAGCTCTCCGGATG
    19321  ---------+---------+---------+---------+---------+---------+  19380
                CCTCCTGTCCTTCAACGCCTAGTAGCCGTACGGCACAAGCCAGGCCTTCGAGAGGCCTAC
    16          S  S  L  F  N  R  I  M  P  M  G  H  E  T  R  F  S  E  P  H  -

                GAACTGGACGGACTCCACCGGCAGCGAACGGTGGCGCAGGCCCATCACGTACCCGTCGTC
    19381  ---------+---------+---------+---------+---------+---------+  19440
                CTTGACCTGCCTGAGGTGGCCGTCGCTTGCCACCGCGTCCGGGTAGTGCATGGGCAGCAG
    16          F  Q  V  S  E  V  P  L  S  R  H  R  L  G  M  V  Y  G  D  D  -

                CGTGGAGCGCCCGGTGACCTCGAGGGACGGCGGGACCGTGCCCTCCGGCACGATCAGTGA
    19441  ---------+---------+---------+---------+---------+---------+  19500
                GCACCTCGCGGGCCACTGGAGCTCCCTGCCGCCCTGGCACGGGAGGCCGTGCTAGTCACT
    16          T  S  R  G  T  V  E  L  S  P  P  V  T  G  E  P  V  I  L  S  -

                GTGGTAGCGGGTCGCGAAGAACCCCGCGGGCAGCCCGGTGAACACTCCGCGCCCGTCGTG
    19501  ---------+---------+---------+---------+---------+---------+  19560
                CACCATCGCCCAGCGCTTCTTGGGGCGCCCGTCGGGCCACTTGTGAGGCGCGGGCAGCAC
    16          H  Y  R  T  A  F  F  G  A  P  L  G  T  F  V  G  R  G  D  H  -

                CGTGATCCGGCTCGTCTTCCCGTGCATGAGATGCCGGGCGGGGACGGTGGCGGCGCCGTA
    19561  ---------+---------+---------+---------+---------+---------+  19620
                GCACTAGGCCGAGCAGAAGGGCACGTACTCTACGGCCCGCCCCTGCCACCGCCGCGGCAT
    16          T  I  R  S  T  K  G  H  M  L  H  R  A  P  V  T  A  A  G  Y  -

                GGCGCGGGCGACGGCCTGATGCCCCAGACAGACCCCGAGCAGCGGGACCCGGCCGGCGAA
    19621  ---------+---------+---------+---------+---------+---------+  19680
                CCGCGCCCGCTGCCGGACTACGGGGTCTGTCTGGGGCTCGTCGCCCTGGGCCGGCCGCTT
    16          A  R  A  V  A  Q  H  G  L  C  V  G  L  L  P  V  R  G  A  F  -

                GGCCTGGACGATCTCGACGTGCCCGGAGGTGTCGGGGTGGCCGGGGCCCGGCCCCAGCAG
    19681  ---------+---------+---------+---------+---------+---------+  19740
                CCGGACCTGCTAGAGCTGCACGGGCCTCCACAGCCCCACCGGCCCCGGGCCGGGGTCGTC
    16          A  Q  V  I  E  V  H  G  S  T  D  P  H  G  P  G  P  G  L  L  -

                GACCGCGTCCGGCCGCATCAGCCCCATCTCGTCCGGGGTCATGAGATGCGACCGCACCAT
    19741  ---------+---------+---------+---------+---------+---------+  19800
                CTGGCGCAGGCCGGCGTAGTCGGGGTAGAGCAGGCCCCAGTACTCTACGCTGGCGTGGTA
    16          V  A  D  P  R  M  L  G  M  E  D  P  T  M  L  H  S  R  V  M  -

                GACGGGCTCCGCGCCGGCGGACATCAGATACTGGCGCAGGATGTCGACGAAGCTGTCGAA
    19801  ---------+---------+---------+---------+---------+---------+  19860
                CTGCCCGAGGCGCGGCCGCCTGTAGTCTATGACCGCGTCCTACAGCTGCTTCGACAGCTT
    16          V  P  E  A  G  A  S  M  L  Y  Q  R  L  I  D  V  F  S  D  F  -

                CGCGTCGACCACCAGGACCCGCGGGGCCTCGGTGCCTGCGCCGGATCCGTCGGGAGACCA
    19861  ---------+---------+---------+---------+---------+---------+  19920
                GCGCAGCTGGTGGTCCTGGGCGCCCCGGAGCCACGGACGCGGCCTAGGCAGCCCTCTGGT
    16          A  D  V  V  L  V  R  P  A  E  T  G  A  G  S  G  D  P  S  W  -

                CAAGCTCACAGCAACTCCTCTCCGGTGACCGCCCAGTGAGTGGCGCTCATCTTGGCCAGC
    19921  ---------+---------+---------+---------+---------+---------+  19980
                GTTCGAGTGTCGTTGAGGAGAGGCCACTGGCGGGTCACTCACCGCGAGTAGAACCGGTCG
```

25

```
16-<        L  S  M  -
15-*           *  L  L  E  E  G  T  V  A  W  H  T  A  S  M  K  A  L  -

         GTCTCGGTCCACTCCGCCCCCGGTTCGGAATCGGCGACGATTCCGGCCGAGGCCCGGGTG
19981    ---------+---------+---------+---------+---------+---------+ 20040
         CAGAGCCAGGTGAGGCGGGGGCCAAGCCTTAGCCGCTGCTAAGGCCGGCTCCGGGCCCAC
15        T  E  T  W  E  A  G  P  E  S  D  A  V  I  G  A  S  A  R  T  -

         CGGTAGACGCCCTCGTGGTGGAAAAGGGTCCGGATGCACAGCGCGAGGTTGGTGTACCCG
20041    ---------+---------+---------+---------+---------+---------+ 20100
         GCCATCTGCGGGAGCACCACCTTTTCCCAGGCCTACGTGTCGCGCTCCAACCACATGGGC
15        R  Y  V  G  E  H  H  F  L  T  R  I  C  L  A  L  N  T  Y  G  -

         CCCACGTCGAGGAGGCCGAGCGCCCCGGCGTACAGGCCGCGGCGGCTGCGTTCGACGGAC
20101    ---------+---------+---------+---------+---------+---------+ 20160
         GGGTGCAGCTCCTCCGGCTCGCGGGGCCGCATGTCCGGCGCCGCCGACGCAAGCTGCCTG
15         G  V  D  L  L  G  L  A  G  A  Y  L  G  R  R  S  R  E  V  S  -

         TCGATGATCTCCATGGCGCGGATCTTCGGCGCGCCCGTCATGGTGCCGGCGGGGAACAGG
20161    ---------+---------+---------+---------+---------+---------+ 20220
         AGCTACTAGAGGTACCGCGCCTAGAAGCCGCGCGGGCAGTACCACGGCCGCCCCTTGTCC
15         E  I  I  E  M  A  R  I  K  P  A  G  T  M  T  G  A  P  F  L  -

         GCGGCGATGGTGTCGAAGGCATCGGTGTCCACCCGCGCCCGGCCGACGACCGTGGAGACC
20221    ---------+---------+---------+---------+---------+---------+ 20280
         CGCCGCTACCACAGCTTCCGTAGCCACAGGTGGGCGCGGGCCGGCTGCTGGCACCTCTGG
15         A  A  I  T  D  F  A  D  T  D  V  R  A  R  G  V  V  T  S  V  -

         AGGTGCAGCACGTGGGAGTAGCCCTCCACGTCCAGCTGGTCGGGTACGTCGAGCGTGTTC
20281    ---------+---------+---------+---------+---------+---------+ 20340
         TCCACGTCGTGCACCCTCATCGGGAGGTGCAGGTCGACCAGCCCATGCAGCTCGCACAAG
15         L  H  L  V  H  S  Y  G  E  V  D  L  Q  D  P  V  D  L  T  N  -

         GGCCGGGCGATCCGTCCGATGTCGTTGCGGCAGAGGTCCACCAGCATGGTGTGCTCGGCG
20341    ---------+---------+---------+---------+---------+---------+ 20400
         CCGGCCCGCTAGGCAGGCTACAGCAACGCCGTCTCCAGGTGGTCGTACCACACGAGCCGC
15         P  R  A  I  R  G  I  D  N  R  C  L  D  V  L  M  T  H  E  A  -

         ATCTCCTTGGGATCCGACCTCAGCCGGACTCCCGCGGCGATGCCGCCGTCCGCGCCGGAC
20401    ---------+---------+---------+---------+---------+---------+ 20460
         TAGAGGAACCCTAGGCTGGAGTCGGCCTGAGGGCGCCGCTACGGCGGCAGGCGCGGCCTG
15         I  E  K  P  D  S  R  L  R  V  G  A  A  I  G  G  D  A  G  S  -

         CGCGGCACCGTGCCCGCGATCGGCCGCATCGTGACCTCGCCGTCCTCGATGCGTACGAAC
20461    ---------+---------+---------+---------+---------+---------+ 20520
         GCGCCGTGGCACGGGCGCTAGCCGGCGTAGCACTGGAGCGGCAGGAGCTACGCATGCTTG
15         R  P  V  T  G  A  I  P  R  M  T  V  E  G  D  E  I  R  V  F  -

         AGCTCGGGGCTGGCGCCGATCAGACGGTGCCCGTCGATGCCCGCCAGATACATGTACGGG
20521    ---------+---------+---------+---------+---------+---------+ 20580
         TCGAGCCCCGACCGCGGCTAGTCTGCCACGGGCAGCTACGGGCGGTCTATGTACATGCCC
15         L  E  P  S  A  G  I  L  R  H  G  D  I  G  A  L  Y  M  Y  P  -

         GAGGCGTTCCGCCCGCGCAGGCGCTGGTAGACGTCCGCGGGGTCGGCCGTCGAGCGGATG
20581    ---------+---------+---------+---------+---------+---------+ 20640
         CTCCGCAAGGCGGGCGCGTCCGCGACCATCTGCAGGCGCCCCAGCCGGCAGCTCGCCTAC
15         S  A  N  R  G  R  L  R  Q  Y  V  D  A  P  D  A  T  S  R  I  -

         GAGAGCTCGTGACCGATCTGCACCTGGTAGATGTCGCCGACGGCGATGTGCTTCAGACAC
20641    ---------+---------+---------+---------+---------+---------+ 20700
         CTCTCGAGCACTGGCTAGACGTGGACCATCTACAGCGGCTGCCGCTACACGAAGTCTGTG
15         S  L  E  H  G  I  Q  V  Q  Y  I  D  G  V  A  I  H  K  L  C  -

         CGCTCGACGTCGTTCGCGAACACTTCGGGGGCGCTGTCGTCGGTGACCGCGGAGGCGGGG
20701    ---------+---------+---------+---------+---------+---------+ 20760
         GCGAGCTGCAGCAAGCGCTTGTGAAGCCCCCGCGACAGCAGCCACTGGCGCCTCCGCCCC
```

26

```
15          R  E  V  D  N  A  F  V  E  P  A  S  D  D  T  V  A  S  A  P -

            AAGCCGTCTGCGGACGGATCGGGCCAGGCCTGCTCCACGTCGGCGAGGAGCCCGGTGACG
      20761 ---------+---------+---------+---------+---------+---------+ 20820
            TTCGGCAGACGCCTGCCTAGCCCGGTCCGGACGAGGTGCAGCCGCTCCTCGGGCCACTGC
15          F  G  D  A  S  P  D  P  W  A  Q  E  V  D  A  L  L  G  T  V -

            GTCTCCGGCGCGAGGCCGGGCCAGTACGGGGACTCGTGGAGCAGCAGTTCGCATCGGCCG
      20821 ---------+---------+---------+---------+---------+---------+ 20880
            CAGAGGCCGCGCTCCGGCCCGGTCATGCCCCTGAGCACCTCGTCGTCAAGCGTAGCCGGC
15          T  E  P  A  L  G  P  W  Y  P  S  E  H  L  L  E  C  R  G -

            GTGGCGAGATCGGTGACCACGCTGCCCCGGTGCAGGACCATGCGTACGTCCGGCAGGCCA
      20881 ---------+---------+---------+---------+---------+---------+ 20940
            CACCGCTCTAGCCACTGGTGCGACGGGGCCACGTCCTGGTACGCATGCAGGCCGTCCGGT
15          T  A  L  D  T  V  V  S  G  R  H  L  V  M  R  V  D  P  L  G -

            GGCCGGTTCTCGATGAGGTGGGGCAGGTCCTCGATGTAGCGGGCCGTGTCGTACCCGAAG
      20941 ---------+---------+---------+---------+---------+---------+ 21000
            CCGGCCAAGAGCTACTCCACCCCGTCCAGGAGCTACATCGCCCGGCACAGCATGGGCTTC
15          P  R  N  E  I  L  H  P  L  D  E  I  Y  R  A  T  D  Y  G  F -

            AACCCGAGGAACCCGAAGCGGAAGCCGGACGCGGACCCCTCGGCGTCGAACATGTCCCGC
      21001 ---------+---------+---------+---------+---------+---------+ 21060
            TTGGGCTCCTTGGGCTTCGCCTTCGGCCTGCGCCTGGGGAGCCGCAGCTTGTACAGGGCG
15          F  G  L  F  G  F  R  F  G  S  A  S  G  E  A  D  F  M  D  R -

            ATGGCCCGCAGCAGCGGCCACAACCCGCCCGCGGTACGCAGCCGCAGCCCCTGGGGGCCG
      21061 ---------+---------+---------+---------+---------+---------+ 21120
            TACCGGGCGTCGTCGCCGGTGTTGGGCGGGCGCCATGCGTCGGCGTCGGGGACCCCCGGC
15          M  A  R  L  L  P  W  L  G  G  A  T  R  L  R  L  G  Q  P  G -

            TCCTCCAGGAGCGCGCCGGCCCGCTCCAGGAGCAGGCCCCGCAGGGCGGGTACGCCCTCG
      21121 ---------+---------+---------+---------+---------+---------+ 21180
            AGGAGGTCCTCGCGCGGCCGGGCGAGGTCCTCGTCCGGGGCGTCCCGCCCATGCGGGAGC
15          D  E  L  L  A  G  A  R  E  L  L  G  R  L  A  P  V  G  E -

            ACGCGCACCACCCGGTCGGTGACCGAGAGCGAGAGCAGCGCGCCGAAGCCGACGAACTGG
      21181 ---------+---------+---------+---------+---------+---------+ 21240
            TGCGCGTGGTGGGCCAGCCACTGGCTCTCGCTCTCGTCGCGCGGCTTCGGCTGCTTGACC
15          V  R  V  V  R  D  T  V  S  L  S  L  L  A  G  F  G  V  F  Q -

            TGCCTGCGGTCGCGGGCCGGGCCGGCCGCGGACTCCAGGAGGTAGACCTCGTCGGGGCCG
      21241 ---------+---------+---------+---------+---------+---------+ 21300
            ACGGACGCCAGCGCCCGGCCCGGCCGGCGCCTGAGGTCCTCCATCTGGAGCAGCCCCGGC
15          H  R  R  D  R  A  P  G  A  A  S  E  L  L  Y  V  E  D  P  G -

            AAGTGCTCGGCCAGCGCGCGGTAGGCGGGCAGGGCGCCCGTCTCCTTCACATCGAGGCGT
      21301 ---------+---------+---------+---------+---------+---------+ 21360
            TTCACGAGCCGGTCGCGCGCCATCCGCCCGTCCCGCGGGCAGAGGAAGTGTAGCTCCGCA
15          F  H  E  A  L  A  R  Y  A  P  L  A  G  T  E  K  V  D  L  R -

            CGTGTCCGCACCCGCACCGGGGCCGAGACCACGCACTGGTCGGTCATCCTGGGTCCTCCC
      21361 ---------+---------+---------+---------+---------+---------+ 21420
            GCACAGGCGTGGGCGTGGCCCCGGCTCTGGTGCGTGACCAGCCAGTAGGACCCAGGAGGG
15-<         R  T  R  V  R  V  P  A  S  V  V  C  Q  D  T  M -

            GGATCACGTGGTGATGGCGTAGCGGTGTGCCACCTGACGGGCGGTCAGCACCGCCCGGTC
      21421 ---------+---------+---------+---------+---------+---------+ 21480
            CCTAGTGCACCACTACCGCATCGCCACACGGTGGACTGCCCGCCAGTCGTGGCGGGCCAG
14-*            *  T  T  I  A  Y  R  H  A  V  Q  R  A  T  L  V  A  R  D -

            GGGGCCGGAGCGGTTGTCGACGACGCGCGCGGCCTTCCAGCTGACGAAGGAGCCGGTGTG
      21481 ---------+---------+---------+---------+---------+---------+ 21540
            CCCCGGCCTCGCCAACAGCTGCTGCGCGCGCCGGAAGGTCGACTGCTTCCTCGGCCACAC
14           P  G  S  R  N  D  V  V  R  A  A  K  W  S  V  F  S  G  T  H -
```

27

```
          GGTCACGGGGTCGAGGTCGGTGTCCACGACGATGCCGGCGTGCGCGCCGGTCCGCTCCCT
    21541 ---------+---------+---------+---------+---------+---------+ 21600
          CCAGTGCCCCAGCTCCAGCCACAGGTGCTGCTACGGCCGCACGCGCGGCCAGGCGAGGGA
   14       T  V  P  D  L  D  T  D  V  V  I  G  A  H  A  G  T  R  E  R  -

          GAGCCGGGCGGCGACGGCCTCGCCGATGCCCTGCCGTTCCCCCTCGGCGCCGGCCAGCAG
    21601 ---------+---------+---------+---------+---------+---------+ 21660
          CTCGGCCCGCCGCTGCCGGAGCGGCTACGGGACGGCAAGGGGGAGCCGCGGCCGGTCGTC
   14       L  R  A  A  V  A  E  G  I  G  Q  R  E  G  E  A  G  A  L  L  -

          GTCCATGCGCACGGTGACGGCGTCGCTGCCGTCGTCCTGCCGGTCGATGACGACCTGGTA
    21661 ---------+---------+---------+---------+---------+---------+ 21720
          CAGGTACGCGTGCCACTGCCGCAGCGACGGCAGCAGGACGGCCAGCTACTGCTGGACCAT
   14       D  M  R  V  T  V  A  D  S  G  D  D  Q  R  D  I  V  V  Q  Y  -

          GCCGAGGCAGCCGCCGACCCCGTCGAGGATCGCGGCCTCCAGCTCGGCGGGCTGGAGGGT
    21721 ---------+---------+---------+---------+---------+---------+ 21780
          CGGCTCCGTCGGCGGCTGGGGCAGCTCCTAGCGCCGGAGGTCGAGCCGCCCGACCTCCCA
   14       G  L  C  G  G  V  G  D  L  I  A  A  E  L  E  A  P  Q  L  T  -

          CACGTCGCCCAGGGGGATGCGGTCCGCGACCCGGCCGATGACCTGGATCCGCGGTCCCGG
    21781 ---------+---------+---------+---------+---------+---------+ 21840
          GTGCAGCGGGTCCCCCTACGCCAGGCGCTGGGCCGGCTACTGGACCTAGGCGCCAGGGCC
   14       V  D  G  L  P  I  R  D  A  V  R  G  I  V  Q  I  R  P  G  P  -

          CAGCGGCTCCCCGGGGCCCGCCGGGAGGATGCGGACCAGGTCCCCGGTGCGGTAGCGGAT
    21841 ---------+---------+---------+---------+---------+---------+ 21900
          GTCGCCGAGGGGCCCCGGGCGGCCCTCCTACGCCTGGTCCAGGGGCCACGCCATCGCCTA
   14       L  P  E  G  P  G  A  P  L  I  R  V  L  D  G  T  R  Y  R  I  -

          CAGTGGTTTGATGCCGTCCACCAGCATGGTGAGGACGAGTTCGCCCTCTCCCGTGTCGCC
    21901 ---------+---------+---------+---------+---------+---------+ 21960
          GTCACCAAACTACGGCAGGTGGTCGTACCACTCCTGCTCAAGCGGGAGAGGGCACAGCGG
   14       L  P  K  I  G  D  V  L  M  T  L  V  L  E  G  E  G  T  D  G  -

          GACCACGGCGCCGGTGTCCGGTTCGACGAGTTCGGTCAAGTAGTTGGGCTGGGCGAGGTG
    21961 ---------+---------+---------+---------+---------+---------+ 22020
          CTGGTGCCGCGGCCACAGGCCAAGCTGCTCAAGCCAGTTCATCAACCCGACCCGCTCCAC
   14       V  V  A  G  T  D  P  E  V  L  E  T  L  Y  N  P  Q  A  L  H  -

          GAGCGCTCCGGTGTCCGCTCCGGTGGCGATGCACAGGGCTTCCTGGGAGCCGTAGAGCGT
    22021 ---------+---------+---------+---------+---------+---------+ 22080
          CTCGCGAGGCCACAGGCGAGGCCACCGCTACGTGTCCCGAAGGACCCTCGGCATCTCGCA
   14       L  A  G  T  D  A  G  T  A  I  C  L  A  E  Q  S  G  Y  L  T  -

          GGGCCGCACGACGGCTTGCGGCCAGAGGGTCGCCACGTTGTCGGCGAACTGCGGGGTGCA
    22081 ---------+---------+---------+---------+---------+---------+ 22140
          CCCGGCGTGCTGCCGAACGCCGGTCTCCCAGCGGTGCAACAGCCGCTTGACGCCCCACGT
   14       P  R  V  V  A  Q  P  W  L  T  A  V  N  D  A  F  Q  P  T  C  -

          GATCTCACCCAGCGTGAGGAAGAGCTTCACGGGAAGCCGGGCCAGGTCGTAGCCGTAGTG
    22141 ---------+---------+---------+---------+---------+---------+ 22200
          CTAGAGTGGGTCGCACTCCTTCTCGAAGTGCCCTTCGGCCCGGTCCAGCATCGGCATCAC
   14       I  E  G  L  T  L  F  L  K  V  P  L  R  A  L  D  Y  G  Y  H  -

          CAGGGCCGCCTTGGCAAGGCTCAGGCACAGCGCCGGAGCACAGACGACGACCTCGACCTC
    22201 ---------+---------+---------+---------+---------+---------+ 22260
          GTCCCGGCGGAACCGTTCCGAGTCCGTGTCGCGGCCTCGTGTCTGCTGCTGGAGCTGGAG
   14       L  A  A  K  A  L  S  L  C  L  A  P  A  C  V  V  V  E  V  E  -

          CAGCTCCTCGATCAGCCGCAGCGCCTTACGGAATCCCACCCTGGGGGACTCGGGCCAGAT
    22261 ---------+---------+---------+---------+---------+---------+ 22320
          GTCGAGGAGCTAGTCGGCGTCGCGGAATGCCTTAGGGTGGGACCCCCTGAGCCCGGTCTA
   14       L  E  E  I  L  R  L  A  K  R  F  G  V  R  P  S  E  P  W  I  -
```

28

```
                       ●                              ●
        CTTGACGTGACAGGCCCCCAGCTCCGCTGCCACCGCGGTGAACACGTCCCCGAACGCGTA
22321   ---------+---------+---------+---------+---------+---------+ 22380
        GAACTGCACTGTCCGGGGGTCGAGGCGACGGTGGCGCCACTTGTGCAGGGGCTTGCGCAT
14        K  V  H  C  A  G  L  E  A  A  V  A  T  F  V  D  G  F  A  Y -

        CAGCTCCGACGGCCCCATCAGGCCCACGACGGGCATCCGCCCCCCGAACCTCGCTTCCAG
22381   ---------+---------+---------+---------+---------+---------+ 22440
        GTCGAGGCTGCCGGGGTAGTCCGGGTGCTGCCCGTAGGCGGGGGGCTTGGAGCGAAGGTC
14        L  E  S  P  G  M  L  G  V  V  P  M  R  G  G  F  R  A  E  L -

        CATGCGGCGCCAGGACTCCCGGACGGCGATGTTGCTGGTCGCGATGTCCTTCTCGCCGCG
22441   ---------+---------+---------+---------+---------+---------+ 22500
        GTACGCCGCGGTCCTGAGGGCCTGCCGCTACAACGACCAGCGCTACAGGAAGAGCGGCGC
14        M  R  R  W  S  E  R  V  A  I  N  S  T  A  I  D  K  E  G  R -

        TGGGCACGGGGTGGCCGCCCCGGTGGTCCCGGTGGTCTCGTAGTAGATGCGTGCTTCGTG
22501   ---------+---------+---------+---------+---------+---------+ 22560
        ACCCGTGCCCCACCGGCGGGGCCACCAGGGCCACCAGAGCATCATCTACGCACGAAGCAC
14        P  C  P  T  A  A  G  T  T  G  T  T  E  Y  Y  I  R  A  E  H -

        CAGCGGGCCCGACAGGACGTCGTGCATCTCCCGCCGCAGGTCGTCCTTGGTGGTGAAGGG
22561   ---------+---------+---------+---------+---------+---------+ 22620
        GTCGCCCGGGCTGTCCTGCAGCACGTAGAGGGCGGCGTCCAGCAGGAACCACCACTTCCC
14        L  P  G  S  L  V  D  H  M  E  R  R  L  D  D  K  T  T  F  P -

        CAGGTCCGCCAGGTTCGCGGGGGTGACGGCCTCGACGTCCACGCCTGCCAGATGGCGGCG
22621   ---------+---------+---------+---------+---------+---------+ 22680
        GTCCAGGCGGTCCAAGCGCCCCCACTGCCGGAGCTGCAGGTGCGGACGGTCTACCGCCGC
14        L  D  A  L  N  A  P  T  V  A  E  V  D  V  G  A  L  H  R  R -

        GTAGAACGGCGAGCGGCGGGTGACGTGGCGCAGTACGGCCGTCAGCCGTTCGCCCTCCCA
22681   ---------+---------+---------+---------+---------+---------+ 22740
        CATCTTGCCGCTCGCCGCCCACTGCACCGCGTCATGCCGGCAGTCGGCAAGCGGGAGGGT
14        Y  F  P  S  R  R  T  V  H  R  L  V  A  T  L  R  E  G  E  W -

        GCGCTCGCGGTCGGCGGCGGTGAGTTCGCCGCGGTAGAACGCGTCGCTCACCTGCCCGTA
22741   ---------+---------+---------+---------+---------+---------+ 22800
        CGCGAGCGCCAGCCGCCGCCACTCAAGCGGCGCCATCTTGCGCAGCGAGTGGACGGGCAT
14        R  E  R  D  A  A  T  L  E  G  R  Y  F  A  D  S  V  Q  G  Y -

        GGCGGACCAGAACTCGCTGTCCGCGTCGGGGTCCAGCGGCCCGGTCCCGCCGGGACCGGG
22801   ---------+---------+---------+---------+---------+---------+ 22860
        CCGCCTGGTCTTGAGCGACAGGCGCAGCCCCAGGTCGCCGGGCCAGGGCGGCCCTGGCCC
14        A  S  W  F  E  S  D  A  D  P  D  L  P  G  T  G  G  P  G  P -

        CCGCCGGCCGTCTCTCACGGCTGTGCCTGGAGTTCGTTGAGCGCGAGGCCGACCCGCTCG
22861   ---------+---------+---------+---------+---------+---------+ 22920
        GGCGGCCGGCAGAGAGTGCCGACACGGACCTCAAGCAACTCGCGCTCCGGCTGGGCGAGC
14-<      R  R  G  D  R  M -
21-*                    *  P  Q  A  Q  L  E  N  L  A  L  G  V  R  E -

        TTGACCTCGTTGGAGGCCAGCACGTCCGAACGGCCGGTGAGCCGACGGTGTTCGTCGAGC
22921   ---------+---------+---------+---------+---------+---------+ 22980
        AACTGGAGCAACCTCCGGTCGTGCAGGCTTGCCGGCCACTCGGCTGCCACAAGCAGCTCG
21        N  V  E  N  S  A  L  V  D  S  R  G  T  L  R  R  H  E  D  L -

        AGTTCGATCATGTCCGTCATCCTCTCGACCAGGCGCGAGACGTTGGTGAGGCCCTCCTCG
22981   ---------+---------+---------+---------+---------+---------+ 23040
        TCAAGCTAGTACAGGCAGTAGGAGAGCTGGTCCGCGCTCTGCAACCACTCCGGGAGGAGC
21        L  E  I  M  D  T  M  R  E  V  L  R  S  V  N  T  L  G  E  E -

        TCCTTGAGCGCGTCGCCCCGGTGCAGCGCGTGCACCGTCGCCGGGAAGCCGCTGCCCACC
23041   ---------+---------+---------+---------+---------+---------+ 23100
        AGGAACTCGCGCAGCGGGGCCACGTCGCGCACGTGGCAGCGGCCCTTCGGCGACGGGTGG
21        D  K  L  A  D  G  R  H  L  A  H  V  T  A  P  F  G  S  G  V -
```

29

```
                    AGGATCATCCGGTTGAGCAGGGCATTGACGGTCAGCTGAGCCCATACCTCGCCGGCGCTG
         23101      ---------+---------+---------+---------+---------+---------+  23160
                    TCCTAGTAGGCCAACTCGTCCCGTAACTGCCAGTCGACTCGGGTATGGAGCGGCCGCGAC
         21           L   I   M   R   N   L   L   A   N   V   T   L   Q   A   W   V   E   G   A   S   -

                    TAGCGGCGGGCGACCGAGATGATCCCCGCGACCTTGTTGCTCAGCGGCCGGTCGAAGCGC
         23161      ---------+---------+---------+---------+---------+---------+  23220
                    ATCGCCGCCCGCTGGCTCTACTAGGGGCGCTGGAACAACGAGTCGCCGGCCAGCTTCGCG
         21           Y   R   R   A   V   S   I   I   G   A   V   K   N   S   L   P   R   D   F   R   -

                    AGATAACCGACTCCGGCACGCTCGATGAAGGTCTGCATGAGGCTGGCCGTGCCGAATCCG
         23221      ---------+---------+---------+---------+---------+---------+  23280
                    TCTATTGGCTGAGGCCGTGCGAGCTACTTCCAGACGTACTCCGACCGGCACGGCTTAGGC
         21           L   Y   G   V   G   A   R   E   I   F   T   Q   M   L   S   A   T   G   F   G   -

                    TGCACGGGCGCCGCGAAGATGATCCCGTCCGCCGCGACCATCTTCGCCACGACCTCGGGC
         23281      ---------+---------+---------+---------+---------+---------+  23340
                    ACGTGCCCGCGGCGCTTCTACTAGGGCAGGCGGCGCTGGTAGAAGCGGTGCTGGAGCCCG
         21           H   V   P   A   A   F   I   I   G   D   A   A   V   M   K   A   V   V   E   P   -

                    ACCCCGTCGGCCAGGGTGCAGGCCACCGGCCTGTCGTTGCAGTCCCCGCAGGGCCCGCAC
         23341      ---------+---------+---------+---------+---------+---------+  23400
                    TGGGGCAGCCGGTCCCACGTCCGGTGGCCGGACAGCAACGTCAGGGGCGTCCCGGGCGTG
         21           V   G   D   A   L   T   C   A   V   P   R   D   N   C   D   G   C   P   G   C   -

                    CGCTCCATCCTGATCGAGCGCAGGTCGACGGCCTCGAAGTCGACGCCGCGGTTCTCTGCT
         23401      ---------+---------+---------+---------+---------+---------+  23460
                    GCGAGGTAGGACTAGCTCGCGTCCAGCTGCCGGAGCTTCAGCTGCGGCGCCAAGAGACGA
         21           R   E   M   R   I   S   R   L   D   V   A   E   F   D   V   G   R   N   E   A   -

                    ACGCGTGCCGCGTGCCGCAGTACGTCGGCGGTGTTGCCGTCACGTTCCGAACCGTTGATC
         23461      ---------+---------+---------+---------+---------+---------+  23520
                    TGCGCACGGCGCACGGCGTCATGCAGCCGCCACAACGGCAGTGCAAGGCTTGGCAACTAG
         21           V   R   A   A   H   R   L   V   D   A   T   N   G   D   R   E   S   G   N   I   -

                    GCGAGGATCTTGAGTTGTGCGCTCACGAGGGGCCTCCTTGGTGAGTCAGGTGCGCTCGGC
         23521      ---------+---------+---------+---------+---------+---------+  23580
                    CGCTCCTAGAACTCAACACGCGAGTGCTCCCCGGAGGAACCACTCAGTCCACGCGAGCCG
         13-*                                                           *   T   R   E   A   -
         21-<         A   L   I   K   L   Q   A   S   M   -

                    GGTCGGCTCGGGGGAACTGTCTGGCCGCCGCTGGTCCGGGAGCCGCAGGGCCGGCTCGGC
         23581      ---------+---------+---------+---------+---------+---------+  23640
                    CCAGCCGAGCCCCCTTGACAGACCGGCGGCGACCAGGCCCTCGGCGTCCCGGCCGAGCCG
         13            T   P   E   P   S   S   D   P   R   R   Q   D   P   L   R   L   A   P   E   A   -

                    GGGGGCGGGAGGAAGACCGCCCCGCGGCGGGCCGCCACGCTCGCCGAACCGGATGAGGGG
         23641      ---------+---------+---------+---------+---------+---------+  23700
                    CCCCCGCCCTCCTTCTGGCGGGGCGCCGCCCGGCGGTGCGAGCGGCTTGGCCTACTCCCC
         13            P   A   P   P   L   G   G   R   P   P   G   G   R   E   G   F   R   I   L   P   -

                    CTTCTCGACGAGATAGAAGCTGATGGTCGCCAGCACGACGCTGATCGAGATCGTGAAGAG
         23701      ---------+---------+---------+---------+---------+---------+  23760
                    GAAGAGCTGCTCTATCTTCGACTACCAGCGGTCGTGCTGCGACTAGCTCTAGCACTTCTC
         13            K   E   V   L   Y   F   S   I   T   A   L   V   V   S   I   S   I   T   F   L   -

                    GAACAGTTCCCAGAACCCCATGTCACCCCGGAATTCCGGCGTTGGCACGGGAGACTTGCC
         23761      ---------+---------+---------+---------+---------+---------+  23820
                    CTTGTCAAGGGTCTTGGGGTACAGTGGGGCCTTAAGGCCGCAACCGTGCCCTCTGAACGG
         13            F   L   E   W   F   G   M   D   G   R   F   E   P   T   P   V   P   S   K   G   -

                    GAAGATGCTGCCGTTCCTGAGCCAGAGGTTGATCACGATCTCGTGCCAGAGGTAGACGCC
         23821      ---------+---------+---------+---------+---------+---------+  23880
                    CTTCTACGACGGCAAGGACTCGGTCTCCAACTAGTGCTAGAGCACGGTCTCCATCTGCGG
         13            F   I   S   G   N   R   L   W   L   N   I   V   I   E   H   W   L   Y   V   G   -

                    GAGGGAGATCTGGCCGAGGAAGAGGATCGGCTTGCTGGTGAAGAGCGCGTCCGAGAACCG
```

30

```
                                                        ●                                        ●

23881  ---------+---------+---------+---------+---------+---------+ 23940
       CTCCCTCTAGACCGGCTCCTTCTCCTAGCCGAACGACCACTTCTCGCGCAGGCTCTTGGC
13        L  S  I  Q  G  L  F  L  I  P  K  S  T  F  L  A  D  S  F  R  -

       GGACTCGGCGCCGGGGACCGTCATCGGTGCCAGGAGCAGCAGGGTGAAGGAGGTCAGGAT
23941  ---------+---------+---------+---------+---------+---------+ 24000
       CCTGAGCCGCGGCCCCTGGCAGTAGCCACGGTCCTCGTCGTCCCACTTCCTCCAGTCCTA
13        S  E  A  G  P  V  T  M  P  A  L  L  L  L  T  F  S  T  L  I  -

       GAAGTGGTCGACGAGCTCCTGGGCCAGGGCCGCGTTGTCGCCCATGCCCGGGATGCCGAT
24001  ---------+---------+---------+---------+---------+---------+ 24060
       CTTCACCAGCTGCTCGAGGACCCGGTCCCGGCGCAACAGCGGGTACGGGCCCTACGGCTA
13        F  H  D  V  L  E  Q  A  L  A  A  N  D  G  M  G  P  I  G  I  -

       GGGCTTGGTGGCGTAGAGGAGGTACAGCGGGATGAGCGGGACCCAGCAGATCAGCGGGCG
24061  ---------+---------+---------+---------+---------+---------+ 24120
       CCCGAACCACCGCATCTCCTCCATGTCGCCCTACTCGCCCTGGGTCGTCTAGTCGCCCGC
13        P  K  T  A  Y  L  L  Y  L  P  I  L  P  V  W  C  I  L  P  R  -

       CCGGATCACGAAACGGTAGAAGCCCGGGGTCCCTGGCGTCGCCTCGGCGTACGCGGAGTA
24121  ---------+---------+---------+---------+---------+---------+ 24180
       GGCCTAGTGCTTTGCCATCTTCGGGCCCCAGGGACCGCAGCGGAGCCGCATGCGCCTCAT
13        R  I  V  F  R  Y  F  G  P  T  G  P  T  A  E  A  Y  A  S  Y  -

       GATGGCCAGTGCCATGCCCGCGGCGAAGCAGCCGGCGTAGTAGGGCGGCCAGTACCACTG
24181  ---------+---------+---------+---------+---------+---------+ 24240
       CTACCGGTCACGGTACGGGCGCCGCTTCGTCGGCCGCATCATCCCGCCGGTCATGGTGAC
13        I  A  L  A  M  G  A  A  F  C  G  A  Y  Y  P  P  W  Y  W  Q  -

       CATCGTCGCGCCGGTGGAGGGGAGGTTGGTGTACGTGACCCAGCCGATGGCCATGACTTC
24241  ---------+---------+---------+---------+---------+---------+ 24300
       GTAGCAGCGCGGCCACCTCCCCTCCAACCACATGCACTGGGTCGGCTACCGGTACTGAAG
13        M  T  A  G  T  S  P  L  N  T  Y  T  V  W  G  I  A  M  V  E  -

       CAGCGCGGCCAGCGGCAGCAGGAGGCGGCGTGCCTTCTGCCCGGGAGTGCTGCCGCCCCG
24301  ---------+---------+---------+---------+---------+---------+ 24360
       GTCGCGCCGGTCGCCGTCGTCCTCCGCCGCACGGAAGACGGGCCCTCACGACGGCGGGGC
13        L  A  A  L  P  L  L  L  R  R  A  K  Q  G  P  T  S  G  G  R  -

       CGCGAGCCGGTGGCCGATCCAGGCGATCAGCGGCAGGGCGAGGTAGAACGTGAACTCGGC
24361  ---------+---------+---------+---------+---------+---------+ 24420
       GCGCTCGGCCACCGGCTAGGTCCGCTAGTCGCCGTCCCGCTCCATCTTGCACTTGAGCCG
13        A  L  R  H  G  I  W  A  I  L  P  L  A  L  Y  F  T  F  E  A  -

       GGGGACCGTCCAGGTGGGCTCGATGCCGTGCATCGGCTGGCCCTCGGGCAGATAGAAGTG
24421  ---------+---------+---------+---------+---------+---------+ 24480
       CCCCTGGCAGGTCCACCCGAGCTACGGCACGTAGCCGACCGGGAGCCCGTCTATCTTCAC
13        P  V  T  W  T  P  E  I  G  H  M  P  Q  G  E  P  L  Y  F  H  -

       CATGAGCAGCACGGGCCGCAGGACGTCGCTGACGCTGTCGATCTCGAACCAGTTGTAGCC
24481  ---------+---------+---------+---------+---------+---------+ 24540
       GTACTCGTCGTGCCCGGCGTCCTGCAGCGACTGCGACAGCTAGAGCTTGGTCAACATCGG
13        M  L  L  V  P  R  L  V  D  S  V  S  D  I  E  F  W  N  Y  G  -

       GGGGATTGCGAAGACGAGCAACAGGTAGTAGGCGGGCAGGATGCGCAGGGCCCGGCGTTT
24541  ---------+---------+---------+---------+---------+---------+ 24600
       CCCCTAACGCTTCTGCTCGTTGTCCATCATCCGCCCGTCCTACGCGTCCCGGGCCGCAAA
13        P  I  A  F  V  L  L  L  Y  Y  A  P  L  I  R  L  A  R  R  K  -

       GAGGAACCGTCCGGTGGCGGGCCGCTTCGTCCCACTGATGGTGACGCGGGCGTAGGGCTT
24601  ---------+---------+---------+---------+---------+---------+ 24660
       CTCCTTGGCAGGCCACCGCCCGGCGAAGCAGGGTGACTACCACTGCGCCCGCATCCCGAA
13        L  F  R  G  T  A  P  R  K  T  G  S  I  T  V  R  A  Y  P  K  -

       GTACAGCATCATTCCGGACAGAGCGAAGAAGGGGGAAGGCATACCCCCAGACCGTCCGCG
24661  ---------+---------+---------+---------+---------+---------+ 24720
```

31

```
                    CATGTCGTAGTAAGGCCTGTCTCGCTTCTTCCCCCTTCCGTATGGGGGTCTGGCAGGCGC
13-<                 Y  L  M  M  G  S  L  A  F  F  P  S  P  M -

                    AGGACGCCCCAGAACGGTTTGCCCGGCTCACCGACGAAGCTGCCCACTCCGGCCTGGAAG
24721   ---------+---------+---------+---------+---------+---------+ 24780
                    TCCTGCGGGGTCTTGCCAAACGGGCCGAGTGGCTGCTTCGACGGGTGAGGCCGGACCTTC

                    GCGACGTGGTAGACGACCACACCCAGCGCGAGGACACCTCGCAGTCCCTCGAACTTCGGT
24781   ---------+---------+---------+---------+---------+---------+ 24840
                    CGCTGCACCATCTGCTGGTGTGGGTCGCGCTCCTGTGGAGCGTCAGGGAGCTTGAAGCCA

                    ATTCGCTTGCTTTTTGCGCCACCTGCGTCGCGAAGGACGTCCCCCATGGAACAGTCCCCT
24841   ---------+---------+---------+---------+---------+---------+ 24900
                    TAAGCGAACGAAAAACGCGGTGGACGCAGCGCTTCCTGCAGGGGGTACCTTGTCAGGGGA

                    TTCCCTTGGCACTTGCTCGTTGACTTCCCGAAATAGTCGGGTCTGCGGAGTGTGAGCCGC
24901   ---------+---------+---------+---------+---------+---------+ 24960
                    AAGGGAACCGTGAACGAGCAACTGAAGGGCTTTATCAGCCCAGACGCCTCACACTCGGCG

                    ATCTCCAATCGTGCTGTTCCGGTGCTCAGGACGACTTGTTTCGGCCTGAGTGGGAAGGCA
24961   ---------+---------+---------+---------+---------+---------+ 25020
                    TAGAGGTTAGCACGACAAGGCCACGAGTCCTGCTGAACAAAGCCGGACTCACCCTTCCGT
12-*                                     *  S  S  K  N  R  G  S  H  S  P -

                    GCCACCCCCGCGCCCCGCCTCGGCCAGACCGGGGGGCCGAGGAGTCCCGTTCCGAGAGGA
25021   ---------+---------+---------+---------+---------+---------+ 25080
                    CGGTGGGGGCGCGGGGCGGAGCCGGTCTGGCCCCCGGCTCCTCAGGGCAAGGCTCTCCT
12                  L  W  G  R  R  G  A  E  A  L  G  P  A  S  S  D  R  E  S  L -

                    TCGGAGTGATCTCCGGCGGCCAGGCGATGCCCACCTCCGGATCCAGCGGATTCAAGCCAT
25081   ---------+---------+---------+---------+---------+---------+ 25140
                    AGCCTCACTAGAGGCCGCCGGTCCGCTACGGGTGGAGGCCTAGGTCGCCTAAGTTCGGTA
12                  I  P  T  I  E  P  P  W  A  I  G  V  E  P  D  L  P  N  L  G -

                    GTTCGAGCCGGGGGTCGTAGGCCGCCGAGCACAGGTAGACGATCACCGCCTCGTCGCTCA
25141   ---------+---------+---------+---------+---------+---------+ 25200
                    CAAGCTCGGCCCCCAGCATCCGGCGGCTCGTGTCCATCTGCTAGTGGCGGAGCAGCGAGT
12                  H  E  L  R  P  D  Y  A  A  S  C  L  Y  V  I  V  A  E  D  S -

                    GCGTGAGGAATCCGAAGCCCAGCCCCGCGGAGACGTACAGCGCCCGTCCGTTCTCCTCGC
25201   ---------+---------+---------+---------+---------+---------+ 25260
                    CGCACTCCTTAGGCTTCGGGTCGGGGCGCCTCTGCATGTCGCGGGCAGGCAAGAGGAGCG
12                  L  T  L  F  G  F  G  L  G  A  S  V  Y  L  A  R  G  N  E  E -

                    CGAGCTCCACGGTCCGCCAGCCGCCGAAGGTGGGCGACCCCACCCGGATGTCGACCACGG
25261   ---------+---------+---------+---------+---------+---------+ 25320
                    GCTCGAGGTGCCAGGCGGTCGGCGGCTTCCACCCGCTGGGGTGGGCCTACAGCTGGTGCC
12                  G  L  E  V  T  R  W  G  G  F  T  P  S  G  V  R  I  D  V  V -

                    CGCCGAACACGCTGCCGCGCAGGCAGCTGAAGTACTTGGCCTGGCCGGGTACGCCCCCGG
25321   ---------+---------+---------+---------+---------+---------+ 25380
                    GCGGCTTGTGCGACGGCGCGTCCGTCGACTTCATGAACCGGACCGGCCCATGCGGGGGCC
12                  A  G  F  V  S  G  R  L  C  S  F  Y  K  A  Q  G  P  V  G  G -

                    CGAAGTGGATGCCCCGCAGCACCCCGTGGGAGGAGATCGCGCAGTTCGCCTGCCGCAGGT
25381   ---------+---------+---------+---------+---------+---------+ 25440
                    GCTTCACCTACGGGGCGTCGTGGGGCACCCTCCTCTAGCGCGTCAAGCGGACGGCGTCCA
12                  A  F  H  I  G  R  L  V  G  H  S  S  I  A  C  N  A  Q  R  L -

                    CGAAGGAGTGGCCTACGGTGCGGCGGAAGGGCTCGCCCTGGAACCACTCGCGAAACGAGC
25441   ---------+---------+---------+---------+---------+---------+ 25500
                    GCTTCCTCACCGGATGCCACGCCGCCTTCCCGAGCGGGACCTTGGTGAGCGCTTTGCTCG
12                  D  F  S  H  G  V  T  R  R  F  P  E  G  Q  F  W  E  R  F  S -

                    CCCGTTCGTCACGGAAGACCTGCTTCTCCTCCGTCCACGCTCCCGAGATCCCGATCGGCT
25501   ---------+---------+---------+---------+---------+---------+ 25560
```

32

```
                 GGGCAAGCAGTGCCTTCTGGACGAAGAGGAGGCAGGTGCGAGGGCTCTAGGGCTAGCCGA
12               G  R  E  D  R  F  V  Q  K  E  E  T  W  A  G  S  I  G  I  P   -

                 TCATCGCTGGCCCCTTCTCTCGACTTCTCTCGACGACTCGCGGGAGGCGGCCGAGGGGTC
      25561      ---------+---------+---------+---------+---------+---------+   25620
                 AGTAGCGACCGGGGAAGAGAGCTGAAGAGAGCTGCTGAGCGCCCTCCGCCGGCTCCCCAG
12-<             K  M  -

                 CGCCGGGCCCGTGGGAACGCCGCAGTCTAGATGCGGCGGCACCGGGGGCAGGGGGGTGCG
      25621      ---------+---------+---------+---------+---------+---------+   25680
                 GCGGCCCGGGCACCCTTGCGGCGTCAGATCTACGCCGCCGTGGCCCCCGTCCCCCCACGC

                 GACGACGTCCGCCCCACCTCAGCACACCGGGAGATGCAGGTCGGTGACGGGCGACGTGAC
      25681      ---------+---------+---------+---------+---------+---------+   25740
                 CTGCTGCAGGCGGGGTGGAGTCGTGTGGCCCTCTACGTCCAGCCACTGCCCGCTGCACTG

                 GATGCAACGGTCCGAGGCCCGGTTGCCCGGACGACGGCCCACAGAGCCATCGGAGCAACG
      25741      ---------+---------+---------+---------+---------+---------+   25800
                 CTACGTTGCCAGGCTCCGGGCCAACGGGCCTGCTGCCGGGTGTCTCGGTAGCCTCGTTGC

                 GAGGCGGACCGCAGATGACCAAGCACGCCCGTGACCGCGCGGTAGTCCTCGGCGCAGGGA
      25801      ---------+---------+---------+---------+---------+---------+   25860
                 CTCCGCCTGGCGTCTACTGGTTCGTGCGGGCACTGGCGCGCCATCAGGAGCCGCGTCCCT
20->                          M  T  K  H  A  R  D  R  A  V  V  L  G  A  G  M -

                 TGGCGGGGCTGCTCGCCGCGCGCGTCCTGTCCGAGACGTACAAGGAAGTGCTGGTGATCG
      25861      ---------+---------+---------+---------+---------+---------+   25920
                 ACCGCCCCGACGAGCGGCGCGCGCAGGACAGGCTCTGCATGTTCCTTCACGACCACTAGC
20               A  G  L  L  A  A  R  V  L  S  E  T  Y  K  E  V  L  V  I  D -

                 ACCGGGACCGGTTGGGCGGCACGGAGCAGCGCCGCGGTGTCCCGCACGGACGCCACGCCC
      25921      ---------+---------+---------+---------+---------+---------+   25980
                 TGGCCCTGGCCAACCCGCCGTGCCTCGTCGCGGCGCCACAGGGCGTGCCTGCGGTGCGGG
20               R  D  R  L  G  G  T  E  Q  R  R  G  V  P  H  G  R  H  A  H -

                 ATGCGCTGCTGGCCAAGGGACAGCAGATCCTCAACGAACTCTTCCCCGGACTCGACACCG
      25981      ---------+---------+---------+---------+---------+---------+   26040
                 TACGCGACGACCGGTTCCCTGTCGTCTAGGAGTTGCTTGAGAAGGGGCCTGAGCTGTGGC
20               A  L  L  A  K  G  Q  Q  I  L  N  E  L  F  P  G  L  D  T  E -

                 AACTCACCTCGGCCGGAATCCCCGCCGGGGACATCGCCGGGAACCTGCGGTGGTACTTCA
      26041      ---------+---------+---------+---------+---------+---------+   26100
                 TTGAGTGGAGCCGGCCTTAGGGGCGGCCCCTGTAGCGGCCCTTGGACGCCACCATGAAGT
20               L  T  S  A  G  I  P  A  G  D  I  A  G  N  L  R  W  Y  F  N -

                 ACGGCCGCCGGCTCCAGCCCTTCGACACCGGGCTGATCAGCGTCTCGGCGACGAGGCCCG
      26101      ---------+---------+---------+---------+---------+---------+   26160
                 TGCCGGCGGCCGAGGTCGGGAAGCTGTGGCCCGACTAGTCGCAGAGCCGCTGCTCCGGGC
20               G  R  R  L  Q  P  F  D  T  G  L  I  S  V  S  A  T  R  P  E -

                 AGCTGGAGTCCCACGTGCGCGCACGGGTCGCCGCGCTGCCACAGGTGAAGATCATGGACG
      26161      ---------+---------+---------+---------+---------+---------+   26220
                 TCGACCTCAGGGTGCACGCGCGTGCCCAGCGGCGCGACGGTGTCCACTTCTAGTACCTGC
20               L  E  S  H  V  R  A  R  V  A  A  L  P  Q  V  K  I  M  D  G -

                 GGTGCGTGATCCGGGGCCTGACCGCCTCGGCCGACCGCAGCCGCGTCACCGGTGTCGAGG
      26221      ---------+---------+---------+---------+---------+---------+   26280
                 CCACGCACTAGGCCCCGGACTGGCGGAGCCGGCTGGCGTCGGCGCAGTGGCCACAGCTCC
20               C  V  I  R  G  L  T  A  S  A  D  R  S  R  V  T  G  V  E  V -

                 TGGTCGACGAGTCGGGTACGGACACCCCGACGCGCCTGGAGGCCGACCTCGTCGTCGACG
      26281      ---------+---------+---------+---------+---------+---------+   26340
                 ACCAGCTGCTCAGCCCATGCCTGTGGGGCTGCGCGGACCTCCGGCTGGAGCAGCAGCTGC
20               V  D  E  S  G  T  D  T  P  T  R  L  E  A  D  L  V  V  D  V -

                 TCACGGGGCGCGGCTCGCGGACTCCCGCCTGGCTGGAGGAGTTCGGATACGAGCGGCCCG
```

33

```
26341 ---------+---------+---------+---------+---------+---------+ 26400
      AGTGCCCCGCGCCGAGCGCCTGAGGGCGGACCGACCTCCTCAAGCCTATGCTCGCCGGGC
20       T  G  R  G  S  R  T  P  A  W  L  E  E  F  G  Y  E  R  P  A -

      CGGAGGACCGCTTCAAGATCGATCTGGCGTACACCACGCGCCACTTCAAGCTCAAGGAAG
26401 ---------+---------+---------+---------+---------+---------+ 26460
      GCCTCCTGGCGAAGTTCTAGCTAGACCGCATGTGGTGCGCGGTGAAGTTCGAGTTCCTTC
20       E  D  R  F  K  I  D  L  A  Y  T  T  R  H  F  K  L  K  E  D -

      ACCCCTACGGCACGGACCTGTCGATCAACCCGGTGGCATCGCCGAGCAACCCGCGCGGCG
26461 ---------+---------+---------+---------+---------+---------+ 26520
      TGGGGATGCCGTGCCTGGACAGCTAGTTGGGCCACCGTAGCGGCTCGTTGGGCGCGCCGC
20       P  Y  G  T  D  L  S  I  N  P  V  A  S  P  S  N  P  R  G  A -

      CGTTCTTCCCCCGGCTCGCGGACGGCAGCTCCCAGCTCTCCCTCACCGGAATCCTCGGCG
26521 ---------+---------+---------+---------+---------+---------+ 26580
      GCAAGAAGGGGGCCGAGCGCCTGCCGTCGAGGGTCGAGAGGGAGTGGCCTTAGGAGCCGC
20       F  F  P  R  L  A  D  G  S  S  Q  L  S  L  T  G  I  L  G  D -

      ACCACCCGCCCACCGACGACGAGGGCTTCCTGGCGTTCGCCAAGTCGCTTGCCGCGCCGG
26581 ---------+---------+---------+---------+---------+---------+ 26640
      TGGTGGGCGGGTGGCTGCTGCTCCCGAAGGACCGCAAGCGGTTCAGCGAACGGCGCGGCC
20       H  P  P  T  D  D  E  G  F  L  A  F  A  K  S  L  A  A  P  E -

      AGATCTACCGGGCCGTCCGCGATGCCGAACCTCTCGACGAACCGGTCACCTTCCGCTTCC
26641 ---------+---------+---------+---------+---------+---------+ 26700
      TCTAGATGGCCCGGCAGGCGCTACGGCTTGGAGAGCTGCTTGGCCAGTGGAAGGCGAAGG
20       I  Y  R  A  V  R  D  A  E  P  L  D  E  P  V  T  F  R  F  P -

      CGGCGAGCGTCCGCCGCCGTTACGAGAGGCTGCGCCGTTTCCCCGGCGGGTTCCTCGTCA
26701 ---------+---------+---------+---------+---------+---------+ 26760
      GCCGCTCGCAGGCGGCGGCAATGCTCTCCGACGCGGCAAAGGGGCCGCCCAAGGAGCAGT
20       A  S  V  R  R  R  Y  E  R  L  R  R  F  P  G  G  F  L  V  M -

      TGGGCGACGGCGTGTGCAGCTTCAACCCCGTCTACGGCCAGGGCATGACGGTCGCCGCCC
26761 ---------+---------+---------+---------+---------+---------+ 26820
      ACCCGCTGCCGCACACGTCGAAGTTGGGGCAGATGCCGGTCCCGTACTGCCAGCGGCGGG
20       G  D  G  V  C  S  F  N  P  V  Y  G  Q  G  M  T  V  A  A  L -

      TGGAGGCCGTGGCGCTGCGGGACCACTTGCGCGACGCCCCGGACCCCGACGCCCTGCGCT
26821 ---------+---------+---------+---------+---------+---------+ 26880
      ACCTCCGGCACCGCGACGCCCTGGTGAACGCGCTGCGGGGCCTGGGGCTGCGGGACGCGA
20       E  A  V  A  L  R  D  H  L  R  D  A  P  D  P  D  A  L  R  F -

      TCTTCCGGCGTATCTCCACGGTCATCGACGTTCCGTGGGACATCGCCGCCGGAGCGGATC
26881 ---------+---------+---------+---------+---------+---------+ 26940
      AGAAGGCCGCATAGAGGTGCCAGTAGCTGCAAGGCACCCTGTAGCGGCGGCCTCGCCTAG
20       F  R  R  I  S  T  V  I  D  V  P  W  D  I  A  A  G  A  D  L -

      TGAACTTCCCCGGGGTGGAGGGCCCCCGCACCATGAAGGTGAAGATGGCCAACGCCTACA
26941 ---------+---------+---------+---------+---------+---------+ 27000
      ACTTGAAGGGGCCCCACCTCCCGGGGGCGTGGTACTTCCACTTCTACCGGTTGCGGATGT
20       N  F  P  G  V  E  G  P  R  T  M  K  V  K  M  A  N  A  Y  M -

      TGGCCCGCCTGCACGCAGCGGCAGCCGTCGACGGCGCGGTGACCGGGGCGTTCTTCCGGG
27001 ---------+---------+---------+---------+---------+---------+ 27060
      ACCGGGCGGACGTGCGTCGCCGTCGGCAGCTGCCGCGCCACTGGCCCCGCAAGAAGGCCC
20       A  R  L  H  A  A  A  A  V  D  G  A  V  T  G  A  F  F  R  V -

      TGGCCGGGCTGGTGGACCCCCCGCAGGCCCTGATGCGCCCCTCCCTCGCCCTGCGGGTCA
27061 ---------+---------+---------+---------+---------+---------+ 27120
      ACCGGCCCGACCACCTGGGGGGCGTCCGGGACTACGCGGGGAGGGAGCGGGACGCCCAGT
20       A  G  L  V  D  P  P  Q  A  L  M  R  P  S  L  A  L  R  V  M -

      TGCGCAACTCCTCGGCGAAGCCGTCGGTCCCTTCGGGCGCCGCCGTATGACCGCGCGGCC
27121 ---------+---------+---------+---------+---------+---------+ 27180
```

34

```
                    ACGCGTTGAGGAGCCGCTTCGGCAGCCAGGGAAGCCCGCGGCGGCATACTGGCGCGCCGG
       20-*          R   N   S   S   A   K   P   S   V   P   S   G   A   A   V   *   -

                    CGTCCGGGGCGGCTGCCGGGGCCAGGAGCCGACATGCGGGTGATGATCACGGTGTTCCCG
       27181        ---------+---------+---------+---------+---------+---------+ 27240
                    GCAGGCCCCGCCGACGGCCCCGGTCCTCGGCTGTACGCCCACTACTAGTGCCACAAGGGC
       19                                             M   R   V   M   I   T   V   F   P   -

                    GCGCGGGCGCACTTCCTGCCGCTGGTGCCCTATGCCTGGGCCCTGCAGAGCGCGGGCCAC
       27241        ---------+---------+---------+---------+---------+---------+ 27300
                    CGCGCCCGCGTGAAGGACGGCGACCACGGGATACGGACCCGGGACGTCTCGCGCCCGGTG
       19           A   R   A   H   F   L   P   L   V   P   Y   A   W   A   L   Q   S   A   G   H   -

                    GAGGTATGTGTCGTGGCGCCCCCGGGCTATCCCACCGGGGTGGCCGACCCCGACTTCCAC
       27301        ---------+---------+---------+---------+---------+---------+ 27360
                    CTCCATACACAGCACCGCGGGGGCCCGATAGGGTGGCCCCACCGGCTGGGGCTGAAGGTG
       19           E   V   C   V   V   A   P   P   G   Y   P   T   G   V   A   D   P   D   F   H   -

                    GAGGCCGTCACCGCGGCCGGCCTGAAGTCGGTGACCTGCGGGCAGCCGCAGCCGCTGGCG
       27361        ---------+---------+---------+---------+---------+---------+ 27420
                    CTCCGGCAGTGGCGCCGGCCGGACTTCAGCCACTGGACGCCCGTCGGCGTCGGCGACCGC
       19           E   A   V   T   A   A   G   L   K   S   V   T   C   G   Q   P   Q   P   L   A   -

                    GTCCACGACCGCGACGACCCCGGCTACGCGGCGATGCTGCCGACCGCGGCGGAGTCGGAG
       27421        ---------+---------+---------+---------+---------+---------+ 27480
                    CAGGTGCTGGCGCTGCTGGGGCCGATGCGCCGCTACGACGGCTGGCGCCGCCTCAGCCTC
       19           V   H   D   R   D   D   P   G   Y   A   A   M   L   P   T   A   A   E   S   E   -

                    CGCTACGTGGCGGCCCTCGGGATCAGCGAGAAGGAGCGCCCCACCTGGGACGTCTTCTAC
       27481        ---------+---------+---------+---------+---------+---------+ 27540
                    GCGATGCACCGCCGGGAGCCCTAGTCGCTCTTCCTCGCGGGGTGGACCCTGCAGAAGATG
       19           R   Y   V   A   A   L   G   I   S   E   K   E   R   P   T   W   D   V   F   Y   -

                    CACTTCACCTTGCTGGCGATCCGCGACTACCATCCGCCGCGGCCGCGGCAGGACGTGGAC
       27541        ---------+---------+---------+---------+---------+---------+ 27600
                    GTGAAGTGGAACGACCGCTAGGCGCTGATGGTAGGCGGCGCCGGCGCCGTCCTGCACCTG
       19           H   F   T   L   L   A   I   R   D   Y   H   P   P   R   P   R   Q   D   V   D   -

                    CAGGTGATCGAGTTCGCCCGGATCTGGCAGCCCGATCTGGTGCTGTGGGACGCCTGGTTC
       27601        ---------+---------+---------+---------+---------+---------+ 27660
                    GTCCACTAGCTCAAGCGGGCCTAGACCGTCGGGCTAGACCACGACACCCTGCGGACCAAG
       19           Q   V   I   E   F   A   R   I   W   Q   P   D   L   V   L   W   D   A   W   F   -

                    CCCTCGGGCGCGATCGCGGCGCGGGTCAGCGGCGCCGCGCACGCGCGGGTGCTCGTAGCC
       27661        ---------+---------+---------+---------+---------+---------+ 27720
                    GGGAGCCCGCGCTAGCGCCGCGCCCAGTCGCCGCGGCGCGTGCGCGCCCACGAGCATCGG
       19           P   S   G   A   I   A   A   R   V   S   G   A   A   H   A   R   V   L   V   A   -

                    CCCGACTACACCGGCTGGGTCACCGAGCGGTTCGCCGCCGCGGGCCCCGCGGCGGGGGCC
       27721        ---------+---------+---------+---------+---------+---------+ 27780
                    GGGCTGATGTGGCCGACCCAGTGGCTCGCCAAGCGGCGGCGCCCGGGGCGCCGCCCCCGG
       19           P   D   Y   T   G   W   V   T   E   R   F   A   A   A   G   P   A   A   G   A   -

                    GACCTCCTGGCCGAGACGATGCGGCCGCTGGCCGAGCGGTACGGCGTGGAGGTCGACGAC
       27781        ---------+---------+---------+---------+---------+---------+ 27840
                    CTGGAGGACCGGCTCTGCTACGCCGGCGACCGGCTCGCCATGCCGCACCTCCAGCTGCTG
       19           D   L   L   A   E   T   M   R   P   L   A   E   R   Y   G   V   E   V   D   D   -

                    GATCTTCTGCTCGGACAGTGGACGGTCAATCCGTTCCCGGCGCCGATGAACCCGCCGACC
       27841        ---------+---------+---------+---------+---------+---------+ 27900
                    CTAGAAGACGAGCCTGTCACCTGCCAGTTAGGCAAGGGCCGCGGCTACTTGGGCGGCTGG
       19           D   L   L   L   G   Q   W   T   V   N   P   F   P   A   P   M   N   P   P   T   -

                    CGGCTCACGAACGTTCCGGTGCGCTACGTGCCCTACACCGGTGCCAGCGTCATGCCCGCG
       27901        ---------+---------+---------+---------+---------+---------+ 27960
                    GCCGAGTGCTTGCAAGGCCACGCGATGCACGGGATGTGGCCACGGTCGCAGTACGGGCGC
```

35

```
                    ●                              ●

19          R  L  T  N  V  P  V  R  Y  V  P  Y  T  G  A  S  V  M  P  A   -

            TGGCTGTACGCGCGGCCGTCGCGGCCGCGGGTGGCGCTGTCGCTCGGAGTGTCCGCGCGG
    27961   ---------+---------+---------+---------+---------+---------+  28020
            ACCGACATGCGCGCCGGCAGCGCCGGCGCCCACCGCGACAGCGAGCCTCACAGGCGCGCC
19          W  L  Y  A  R  P  S  R  P  R  V  A  L  S  L  G  V  S  A  R   -

            GCGTTCCTCAAGGGTGACTGGGGGCGTACCGCCAAACTGCTGGAAGCGGTCGCGGAGCTG
    28021   ---------+---------+---------+---------+---------+---------+  28080
            CGCAAGGAGTTCCCACTGACCCCCGCATGGCGGTTTGACGACCTTCGCCAGCGCCTCGAC
19          A  F  L  K  G  D  W  G  R  T  A  K  L  L  E  A  V  A  E  L   -

            GACATCGAGGTGATCGCCACGCTCAACGACAACCAACTGGCGGAGAGCGGGCCGCTGCCG
    28081   ---------+---------+---------+---------+---------+---------+  28140
            CTGTAGCTCCACTAGCGGTGCGAGTTGCTGTTGGTTGACCGCCTCTCGCCCGGCGACGGC
19          D  I  E  V  I  A  T  L  N  D  N  Q  L  A  E  S  G  P  L  P   -

            GACAACGTCCACACCCTCGACTACGTACCGCTCGACCAGTTGCTGCCCACCTGCTCGGCC
    28141   ---------+---------+---------+---------+---------+---------+  28200
            CTGTTGCAGGTGTGGGAGCTGATGCATGGCGAGCTGGTCAACGACGGGTGGACGAGCCGG
19          D  N  V  H  T  L  D  Y  V  P  L  D  Q  L  L  P  T  C  S  A   -

            GTCATCCACCACGGATCGACGGGCACCTTCGCCGCGGCGAGCGCGGCCGGGCTGCCCCAG
    28201   ---------+---------+---------+---------+---------+---------+  28260
            CAGTAGGTGGTGCCTAGCTGCCCGTGGAAGCGGCGCCGCTCGCGCCGGCCCGACGGGGTC
19          V  I  H  H  G  S  T  G  T  F  A  A  A  S  A  A  G  L  P  Q   -

            GTGGTCTGCGACACCGACGAGCCCCTCCTGCTCTTCGGCGAGGACACCCCCGACGGCATC
    28261   ---------+---------+---------+---------+---------+---------+  28320
            CACCAGACGCTGTGGCTGCTCGGGGAGGACGAGAAGCCGCTCCTGTGGGGGCTGCCGTAG
19          V  V  C  D  T  D  E  P  L  L  L  F  G  E  D  T  P  D  G  I   -

            GCGTGGGACTTCACCTGCCAGAAGCAGCTCACCGCGACGCTCACCTCCCGCGTGGTCACC
    28321   ---------+---------+---------+---------+---------+---------+  28380
            CGCACCCTGAAGTGGACGGTCTTCGTCGAGTGGCGCTGCGAGTGGAGGGCGCACCAGTGG
19          A  W  D  F  T  C  Q  K  Q  L  T  A  T  L  T  S  R  V  V  T   -

            GACTACGGGGCGGGGGTGCGCGTCGACCACCAGAAGCAGTCCGCCGGACAGATCCGTGAG
    28381   ---------+---------+---------+---------+---------+---------+  28440
            CTGATGCCCCGCCCCCACGCGCAGCTGGTGGTCTTCGTCAGGCGGCCTGTCTAGGCACTC
19          D  Y  G  A  G  V  R  V  D  H  Q  K  Q  S  A  G  Q  I  R  E   -

            CAACTACGCAGGGTGCTCACCGAACCTTCCTTCCGCGAGGGCGCTCGACGGATCCGGGAA
    28441   ---------+---------+---------+---------+---------+---------+  28500
            GTTGATGCGTCCCACGAGTGGCTTGGAAGGAAGGCGCTCCCGCGAGCTGCCTAGGCCCTT
19          Q  L  R  R  V  L  T  E  P  S  F  R  E  G  A  R  R  I  R  E   -

            GACCGGAATTCCGCCCCCAGCCCGGTCGAACTCGTATCGCTCCTGGTAGAACTGACGAAG
    28501   ---------+---------+---------+---------+---------+---------+  28560
            CTGGCCTTAAGGCGGGGGTCGGGCCAGCTTGAGCATAGCGAGGACCATCTTGACTGCTTC
19          D  R  N  S  A  P  S  P  V  E  L  V  S  L  L  V  E  L  T  K   -

            CGTCATCGCCGTGACAAGGAGGCGGACCGATGAGGATGCTGGTGACGGGCGGAGCGGGTT
    28561   ---------+---------+---------+---------+---------+---------+  28620
            GCAGTAGCGGCACTGTTCCTCCGCCTGGCTACTCCTACGACCACTGCCCGCCTCGCCCAA
19-*        R  H  R  R  D  K  E  A  D  R  *  -
1->                                      M  R  M  L  V  T  G  G  A  G  F  -

            TCATCGGCTCGCAGTTCGTGCGGGCCACACTGCACGGCGAGCTGCCGGGGTTCCGAGGACG
    28621   ---------+---------+---------+---------+---------+---------+  28680
            AGTAGCCGAGCGTCAAGCACGCCCGGTGTGACGTGCCGCTCGACGGCCCAAGGCTCCTGC
1              I  G  S  Q  F  V  R  A  T  L  H  G  E  L  P  G  S  E  D  A   -

            CCCGGGTGACGGTCCTGGACAAGCTGACGTACTCCGGCAATCCGGCCAACCTCACCTCCG
    28681   ---------+---------+---------+---------+---------+---------+  28740
            GGGCCCACTGCCAGGACCTGTTCGACTGCATGAGGCCGTTAGGCCGGTTGGAGTGGAGGC
```

36

```
                    R   V   T   V   L   D   K   L   T   Y   S   G   N   P   A   N   L   T   S   V  -
         1

              TCGCGGCCCATCCGCGGTACACCTTCGTCCAGGGCGACACCGTCGACCCGCGCGTCGTCG
     28741    ---------+---------+---------+---------+---------+---------+   28800
              AGCGCCGGGTAGGCGCCATGTGGAAGCAGGTCCCGCTGTGGCAGCTGGGCGCGCAGCAGC
         1        A   A   H   P   R   Y   T   F   V   Q   G   D   T   V   D   P   R   V   V   D  -

              ACGAGGTGGTCGCCGGCCACGACGTCATCGTCCACTTCGCGGCGGAGTCGCACGTGGACC
     28801    ---------+---------+---------+---------+---------+---------+   28860
              TGCTCCACCAGCGGCCGGTGCTGCAGTAGCAGGTGAAGCGCCGCCTCAGCGTGCACCTGG
         1          E   V   V   A   G   H   D   V   I   V   H   F   A   A   E   S   H   V   D   R  -

              GCTCGATCGACACCGCCACCCGGTTCGTCACGACCAACGTGCTCGGGACCCAGACGCTGC
     28861    ---------+---------+---------+---------+---------+---------+   28920
              CGAGCTAGCTGTGGCGGTGGGCCAAGCAGTGCTGGTTGCACGAGCCCTGGGTCTGCGACG
         1          S   I   D   T   A   T   R   F   V   T   T   N   V   L   G   T   Q   T   L   L  -

              TGGAAGCGGCTCTCCGGCACGGGGTCGGCCGGTTCGTGCACGTGTCGACCGACGAGGTCT
     28921    ---------+---------+---------+---------+---------+---------+   28980
              ACCTTCGCCGAGAGGCCGTGCCCCAGCCGGCCAAGCACGTGCACAGCTGGCTGCTCCAGA
         1          E   A   A   L   R   H   G   V   G   R   F   V   H   V   S   T   D   E   V   Y  -

              ACGGGTCGATCGCCTCCGGCTCATGGACCGAGGACACCCCGCTCGCCCCCAACGTCCCCT
     28981    ---------+---------+---------+---------+---------+---------+   29040
              TGCCCAGCTAGCGGAGGCCGAGTACCTGGCTCCTGTGGGGCGAGCGGGGGTTGCAGGGGA
         1          G   S   I   A   S   G   S   W   T   E   D   T   P   L   A   P   N   V   P   Y  -

              ACGCGGCGTCGAAGGCGGGTTCGGACCTGATGGCGCTCGCCTGGCACCGCACCCGGGGCC
     29041    ---------+---------+---------+---------+---------+---------+   29100
              TGCGCCGCAGCTTCCGCCCAAGCCTGGACTACCGCGAGCGGACCGTGGCGTGGGCCCCGG
         1          A   A   S   K   A   G   S   D   L   M   A   L   A   W   H   R   T   R   G   L  -

              TGGACGTCGTCGTCACCCGGTGCACCAACAACTACGGTCCCTACCAGTACCCCGAGAAGG
     29101    ---------+---------+---------+---------+---------+---------+   29160
              ACCTGCAGCAGCAGTGGGCCACGTGGTTGTTGATGCCAGGGATGGTCATGGGGCTCTTCC
         1          D   V   V   V   T   R   C   T   N   N   Y   G   P   Y   Q   Y   P   E   K   V  -

              TGATCCCGCTCTTCGTCACCAACATCCTCGACGGCTTGCGGGTGCCCCTGTACGGGGACG
     29161    ---------+---------+---------+---------+---------+---------+   29220
              ACTAGGGCGAGAAGCAGTGGTTGTAGGAGCTGCCGAACGCCCACGGGGACATGCCCCTGC
         1          I   P   L   F   V   T   N   I   L   D   G   L   R   V   P   L   Y   G   D   G  -

              GCGCCCACCGCCGGGACTGGCTGCACGTGTCCGACCACTGCCGGGCCATCCAGATGGTCA
     29221    ---------+---------+---------+---------+---------+---------+   29280
              CGCGGGTGGCGGCCCTGACCGACGTGCACAGGCTGGTGACGGCCCGGTAGGTCTACCAGT
         1          A   H   R   R   D   W   L   H   V   S   D   H   C   R   A   I   Q   M   V   M  -

              TGAACTCCGGCCGGGCCGGGGAGGTCTACCACATCGGCGGCGGCACCGAACTCTCCAACG
     29281    ---------+---------+---------+---------+---------+---------+   29340
              ACTTGAGGCCGGCCCGGCCCCTCCAGATGGTGTAGCCGCCGCCGTGGCTTGAGAGGTTGC
         1          N   S   G   R   A   G   E   V   Y   H   I   G   G   G   T   E   L   S   N   E  -

              AGGAACTCACCGGCCTGTTGCTCACGGCGTGCGGCACCGACTGGTCCTGCGTGGACCGGG
     29341    ---------+---------+---------+---------+---------+---------+   29400
              TCCTTGAGTGGCCGGACAACGAGTGCCGCACGCCGTGGCTGACCAGGACGCACCTGGCCC
         1          E   L   T   G   L   L   L   T   A   C   G   T   D   W   S   C   V   D   R   V  -

              TGGCCGACCGGCAGGGGCACGACCGCCGCTACTCGCTCGACATCACGAAGATCCGGCAGG
     29401    ---------+---------+---------+---------+---------+---------+   29460
              ACCGGCTGGCCGTCCCCGTGCTGGCGGCGATGAGCGAGCTGTAGTGCTTCTAGGCCGTCC
         1          A   D   R   Q   G   H   D   R   R   Y   S   L   D   I   T   K   I   R   Q   E  -

              AACTGGGCTACGAGCCCCTGGTCGCCTTCGAGGACGGCCTGGCCGCGACGGTGAAGTGGT
     29461    ---------+---------+---------+---------+---------+---------+   29520
              TTGACCCGATGCTCGGGGACCAGCGGAAGCTCCTGCCGGACCGGCGCTGCCACTTCACCA
         1          L   G   Y   E   P   L   V   A   F   E   D   G   L   A   A   T   V   K   W   Y  -
```

37

```
         ACCACGAGAACCGTTCGTGGTGGCAGCCGCTGAAGGAAGCGGCCGGCCTCCTGGACGCCG
29521    ---------+---------+---------+---------+---------+---------+    29580
         TGGTGCTCTTGGCAAGCACCACCGTCGGCGACTTCCTTCGCCGGCCGGAGGACCTGCGGC
1          H  E  N  R  S  W  W  Q  P  L  K  E  A  A  G  L  L  D  A  V  -

         TCGGCTGACGGCAGCCACCGCTAGGAACACCCCAGGAAAGGAGCCACCTCCGTGACAGCA
29581    ---------+---------+---------+---------+---------+---------+    29640
         AGCCGACTGCCGTCGGTGGCGATCCTTGTGGGGTCCTTTCCTCGGTGGAGGCACTGTCGT
2->                                                          M  T  A  -
1-*          G  *  -

         GTCAAGGAGCCGACGTCCCGCGCAGGACGGCGGGAGTGGATCGCTCTCGTCGTCCTCTCC
29641    ---------+---------+---------+---------+---------+---------+    29700
         CAGTTCCTCGGCTGCAGGGCGCGTCCTGCCGCCCTCACCTAGCGAGAGCAGCAGGAGAGG
2          V  K  E  P  T  S  R  A  G  R  R  E  W  I  A  L  V  V  L  S  -

         TTGCCCACGATGCTGTTGATGCTGGACATCAACGTCCTCATGCTGGCCTTGCCGCAGTTG
29701    ---------+---------+---------+---------+---------+---------+    29760
         AACGGGTGCTACGACAACTACGACCTGTAGTTGCAGGAGTACGACCGGAACGGCGTCAAC
2          L  P  T  M  L  L  M  L  D  I  N  V  L  M  L  A  L  P  Q  L  -

         AGCGAGGATCTCGGCGCGAGCAGCACGCAACAGCTGTGGATCACCGACATCTACGGATTC
29761    ---------+---------+---------+---------+---------+---------+    29820
         TCGCTCCTAGAGCCGCGCTCGTCGTGCGTTGTCGACACCTAGTGGCTGTAGATGCCTAAG
2          S  E  D  L  G  A  S  S  T  Q  Q  L  W  I  T  D  I  Y  G  F  -

         GCGATCGCCGGCTTCCTGGTGACCATGGGCACCCTCGGCGACCGGATCGGCCGCCGCAGG
29821    ---------+---------+---------+---------+---------+---------+    29880
         CGCTAGCGGCCGAAGGACCACTGGTACCCGTGGGAGCCGCTGGCCTAGCCGGCGGCGTCC
2          A  I  A  G  F  L  V  T  M  G  T  L  G  D  R  I  G  R  R  R  -

         CTCCTGCTCGGGGGCGCGGCCGTCTTCGCGGTCGTGTCCGTCGTCGCCGCGTTCTCCGAC
29881    ---------+---------+---------+---------+---------+---------+    29940
         GAGGACGAGCCCCCGCGCCGGCAGAAGCGCCAGCACAGGCAGCAGCGGCGCAAGAGGCTG
2          L  L  L  G  G  A  A  V  F  A  V  V  S  V  V  A  A  F  S  D  -

         AGCGCGGCGATGCTCGTCGTCAGCCGCGCCGTGCTCGGCGTCGCCGGGGCCACGGTGATG
29941    ---------+---------+---------+---------+---------+---------+    30000
         TCGCGCCGCTACGAGCAGCAGTCGGCGCGGCACGAGCCGCAGCGGCCCCGGTGCCACTAC
2          S  A  A  M  L  V  V  S  R  A  V  L  G  V  A  G  A  T  V  M  -

         CCCTCGACGCTCGCGCTCATCAGCAACATGTTCGAGGACCCCAAGGAGCGGGGCACCGCC
30001    ---------+---------+---------+---------+---------+---------+    30060
         GGGAGCTGCGAGCGCGAGTAGTCGTTGTACAAGCTCCTGGGGTTCCTCGCCCCGTGGCGG
2          P  S  T  L  A  L  I  S  N  M  F  E  D  P  K  E  R  G  T  A  -

         ATCGCCATGTGGGCGAGCGCCATGATGGCCGGAGTCGCCCTCGGGCCCGCCGTCGGCGGC
30061    ---------+---------+---------+---------+---------+---------+    30120
         TAGCGGTACACCCGCTCGCGGTACTACCGGCCTCAGCGGGAGCCCGGGCGGCAGCCGCCG
2          I  A  M  W  A  S  A  M  M  A  G  V  A  L  G  P  A  V  G  G  -

         CTGGTCCTCGCCGCGTTCTGGTGGGGATCGGTGTTCCTCATCGCCGTTCCGGTGATGCTG
30121    ---------+---------+---------+---------+---------+---------+    30180
         GACCAGGAGCGGCGCAAGACCACCCCTAGCCACAAGGAGTAGCGGCAAGGCCACTACGAC
2          L  V  L  A  A  F  W  W  G  S  V  F  L  I  A  V  P  V  M  L  -

         CTGGTGGTGGTCACCGGCCCCGTGCTGCTCACCGAGTCCCGCGACCCGGACGCCGGACGG
30181    ---------+---------+---------+---------+---------+---------+    30240
         GACCACCACCAGTGGCCGGGGCACGACGAGTGGCTCAGGGCGCTGGGCCTGCGGCCTGCC
2          L  V  V  V  T  G  P  V  L  L  T  E  S  R  D  P  D  A  G  R  -

         CTGGACCTGCTGAGCGCGGGGCTCTCCCTCGCGACCGTGCTGCCGGTGATCTACGGACTG
30241    ---------+---------+---------+---------+---------+---------+    30300
         GACCTGGACGACTCGCGCCCCGAGAGGGAGCGCTGGCACGACGGCCACTAGATGCCTGAC
2          L  D  L  L  S  A  G  L  S  L  A  T  V  L  P  V  I  Y  G  L  -
```

38

```
          AAGGAGCTGGCCCGGACCGGGTGGGACCCGCTCGCCGCCGGCGCGGTGGTCCTCGGCGTG
30301     ---------+---------+---------+---------+---------+---------+   30360
          TTCCTCGACCGGGCCTGGCCCACCCTGGGCGAGCGGCGGCCGCGCCACCAGGAGCCGCAC
2          K  E  L  A  R  T  G  W  D  P  L  A  A  G  A  V  V  L  G  V   -

          ATCTTCGGCGCGCTGTTCGTCCAGCGCCAGCGGCGGTTGGCCGACCCCATGCTGGACCTC
30361     ---------+---------+---------+---------+---------+---------+   30420
          TAGAAGCCGCGCGACAAGCAGGTCGCGGTCGCCGCCAACCGGCTGGGGTACGACCTGGAG
2          I  F  G  A  L  F  V  Q  R  Q  R  R  L  A  D  P  M  L  D  L   -

          GGCCTCTTCGCCGACCGCACCCTGCGGGCGGGTCTGACGGTCAGTCTGGTCAACGCCGTC
30421     ---------+---------+---------+---------+---------+---------+   30480
          CCGGAGAAGCGGCTGGCGTGGGACGCCCGCCCAGACTGCCAGTCAGACCAGTTGCGGCAG
2          G  L  F  A  D  R  T  L  R  A  G  L  T  V  S  L  V  N  A  V   -

          ATCATGGGCGGGACCGGACTGATGGTCGCCCTGTACCTCCAGACGATCGCCGGTCACTCC
30481     ---------+---------+---------+---------+---------+---------+   30540
          TAGTACCCGCCCTGGCCTGACTACCAGCGGGACATGGAGGTCTGCTAGCGGCCAGTGAGG
2          I  M  G  G  T  G  L  M  V  A  L  Y  L  Q  T  I  A  G  H  S   -

          CCGTTGGCCGCCGGGCTGTGGCTGCTGATCCCGGCCTGCATGCTCGTCGTGGGCGTACAG
30541     ---------+---------+---------+---------+---------+---------+   30600
          GGCAACCGGCGGCCCGACACCGACGACTAGGGCCGGACGTACGAGCAGCACCCGCATGTC
2          P  L  A  A  G  L  W  L  L  I  P  A  C  M  L  V  V  G  V  Q   -

          CTGTCGAACCTGCTGGCCCAGCGGATGCCCCCTTCCCGGGTGCTGCTGGGGGGACTGCTG
30601     ---------+---------+---------+---------+---------+---------+   30660
          GACAGCTTGGACGACCGGGTCGCCTACGGGGGAAGGGCCCACGACGACCCCCCTGACGAC
2          L  S  N  L  L  A  Q  R  M  P  P  S  R  V  L  L  G  G  L  L   -

          ATCGCGGCCGTCGGACAGCTCCTGATCACCCAGGTGGACACCGAGGACACCGCCCTCCTC
30661     ---------+---------+---------+---------+---------+---------+   30720
          TAGCGCCGGCAGCCTGTCGAGGACTAGTGGGTCCACCTGTGGCTCCTGTGGCGGGAGGAG
2          I  A  A  V  G  Q  L  L  I  T  Q  V  D  T  E  D  T  A  L  L   -

          ATCGCGGCCACCACCCTGATCTACTTCGGCGCCTCACCGGTGGGGCCGATCACCACGGGC
30721     ---------+---------+---------+---------+---------+---------+   30780
          TAGCGCCGGTGGTGGGACTAGATGAAGCCGCGGAGTGGCCACCCCGGCTAGTGGTGCCCG
2          I  A  A  T  T  L  I  Y  F  G  A  S  P  V  G  P  I  T  T  G   -

          GCGATCATGGGAGCCGCGCCCCCGGAGAAGGCGGGTGCCGCCTCGTCGCTGTCCGCCACC
30781     ---------+---------+---------+---------+---------+---------+   30840
          CGCTAGTACCCTCGGCGCGGGGGCCTCTTCCGCCCACGGCGGAGCAGCGACAGGCGGTGG
2          A  I  M  G  A  A  P  P  E  K  A  G  A  A  S  S  L  S  A  T   -

          GGCGGCGAGTTCGGAGTGGCGCTCGGCATCGCGGGCCTGGGGAGTCTGGGCACCGTCGTG
30841     ---------+---------+---------+---------+---------+---------+   30900
          CCGCCGCTCAAGCCTCACCGCGAGCCGTAGCGCCCGGACCCCTCAGACCCGTGGCAGCAC
2          G  G  E  F  G  V  A  L  G  I  A  G  L  G  S  L  G  T  V  V   -

          TACAGCGCCGGGGTCGAGGTGCCGGACGCGGCCGGGCCCGCCGACGCCGACGCCGCGCAG
30901     ---------+---------+---------+---------+---------+---------+   30960
          ATGTCGCGGCCCCAGCTCCACGGCCTGCGCCGGCCCGGGCGGCTGCGGCTGCGGCGCGTC
2          Y  S  A  G  V  E  V  P  D  A  A  G  P  A  D  A  D  A  A  Q   -

          GAGAGCATCGCCGGCGCCCTGCACACGGCCGGTCAGCTGGCACCGGGCAGCGCCGACGCC
30961     ---------+---------+---------+---------+---------+---------+   31020
          CTCTCGTAGCGGCCGCGGGACGTGTGCCGGCCAGTCGACCGTGGCCCGTCGCGGCTGCGG
2          E  S  I  A  G  A  L  H  T  A  G  Q  L  A  P  G  S  A  D  A   -

          CTGCTGGACTCCGCGCGCGCGGCCTTCACCAGCGGCGTGCAGTCCGTCGCCGCCGTCTGC
31021     ---------+---------+---------+---------+---------+---------+   31080
          GACGACCTGAGGCGCGCGCGCCGGAAGTGGTCGCCGCACGTCAGGCAGCGGCGGCAGACG
2          L  L  D  S  A  R  A  A  F  T  S  G  V  Q  S  V  A  A  V  C   -
```

39

```
                    GCCGTGTTCTCCCTGGCGCTCGCCGTCCTCATCGGCACCCGGCTGCGGGACATTTCCGCG
      31081         ---------+---------+---------+---------+---------+---------+  31140
                    CGGCACAAGAGGGACCGCGAGCGGCAGGAGTAGCCGTGGGCCGACGCCCTGTAAAGGCGC
      2              A  V  F  S  L  A  L  A  V  L  I  G  T  R  L  R  D  I  S  A   -

                    ATGGACCACGGGCACGGCGAGGAACCGGCCGAGAACGACGCTCAACCGGCCACATGAGCG
      31141         ---------+---------+---------+---------+---------+---------+  31200
                    TACCTGGTGCCCGTGCCGCTCCTTGGCCGGCTCTTGCTGCGAGTTGGCCGGTGTACTCGC
      2-*           M  D  H  G  H  G  E  E  P  A  E  N  D  A  Q  P  A  T  *  -

                    CACTTCCGGAGATGCAACGGCCGCCGTCGAGGTATGAGGATCACCTTCCGGGGTGCACCT
      31201         ---------+---------+---------+---------+---------+---------+  31260
                    GTGAAGGCCTCTACGTTGCCGGCGGCAGCTCCATACTCCTAGTGGAAGGCCCCACGTGGA

                    GCACGGCAACGGAGGCGTAGTGGAGTACTGGAACAGCACGGCGGAGACCATGCCCCGCCA
      31261         ---------+---------+---------+---------+---------+---------+  31320
                    CGTGCCGTTGCCTCCGCATCACCTCATGACCTTGTCGTGCCGCCTCTGGTACGGGGCGGT
      3->                        M  E  Y  W  N  S  T  A  E  T  M  P  R  Q   -

                    GGAACTCGAACAGTGGAAGTGGCGCAGGCTCCAGGCCGCCATGGACCACGCCAGAAGGCT
      31321         ---------+---------+---------+---------+---------+---------+  31380
                    CCTTGAGCTTGTCACCTTCACCGCGTCCGAGGTCCGGCGGTACCTGGTGCGGTCTTCCGA
      3              E  L  E  Q  W  K  W  R  R  L  Q  A  A  M  D  H  A  R  R  L   -

                    TTCGCCCTTCTGGCGGGAACGACTCCCCGAGAACATCACCTCCATGGCGGACTACGCGGC
      31381         ---------+---------+---------+---------+---------+---------+  31440
                    AAGCGGGAAGACCGCCCTTGCTGAGGGGCTCTTGTAGTGGAGGTACCGCCTGATGCGCCG
      3              S  P  F  W  R  E  R  L  P  E  N  I  T  S  M  A  D  Y  A  A   -

                    GCGGGTGCCTCTCCTGCGCAAGGCCGACCTCCTCGCCGCGGAAGCCGCGTCTCCCCCTTA
      31441         ---------+---------+---------+---------+---------+---------+  31500
                    CGCCCACGGAGAGGACGCGTTCCGGCTGGAGGAGCGGCGCCTTCGGCGCAGAGGGGGAAT
      3              R  V  P  L  L  R  K  A  D  L  L  A  A  E  A  A  S  P  P  Y   -

                    CGGCACCTGGCCCTCGCTGGATCCGGCGCTCGGAGTGCGCCATCACCAGACCAGCGGCAC
      31501         ---------+---------+---------+---------+---------+---------+  31560
                    GCCGTGGACCGGGAGCGACCTAGGCCGCGAGCCTCACGCGGTAGTGGTCTGGTCGCCGTG
      3              G  T  W  P  S  L  D  P  A  L  G  V  R  H  H  Q  T  S  G  T   -

                    CAGCGGTAACCCCCCCATCCGGACGTTCGACACCGAACGCGACTGGGCCTGGTGCGTGGA
      31561         ---------+---------+---------+---------+---------+---------+  31620
                    GTCGCCATTGGGGGGGTAGGCCTGCAAGCTGTGGCTTGCGCTGACCCGGACCACGCACCT
      3              S  G  N  P  P  I  R  T  F  D  T  E  R  D  W  A  W  C  V  D   -

                    CACGTTCTGCACGGCGCTCCACAGCATGGGCGTGCGCCCGCACCACAAGGGTCTGGTGGC
      31621         ---------+---------+---------+---------+---------+---------+  31680
                    GTGCAAGACGTGCCGCGAGGTGTCGTACCCGCACGCGGGCGTGGTGTTCCCAGACCACCG
      3              T  F  C  T  A  L  H  S  M  G  V  R  P  H  H  K  G  L  V  A   -

                    GTTCGGCTACGGGCTGTTCGCCGGTTTCTGGGGCATGCACTACGGCCTCGAGCGCATGGG
      31681         ---------+---------+---------+---------+---------+---------+  31740
                    CAAGCCGATGCCCGACAAGCGGCCAAAGACCCCGTACGTGATGCCGGAGCTCGCGTACCC
      3              F  G  Y  G  L  F  A  G  F  W  G  M  H  Y  G  L  E  R  M  G   -

                    CGCCACGGTCATCCCGGCCGGCGGCCTCGACTCCCGCTCCCGGGTACGGCTGCTGGTCGA
      31741         ---------+---------+---------+---------+---------+---------+  31800
                    GCGGTGCCAGTAGGGCCGGCCGCCGGAGCTGAGGGCGAGGGCCCATGCCGACGACCAGCT
      3              A  T  V  I  P  A  G  G  L  D  S  R  S  R  V  R  L  L  V  D   -

                    CTACCAGATCGAGGTGCTCGGCCTCACACCGAGCTATGCGATGCGGCTGATCGAGACGGC
      31801         ---------+---------+---------+---------+---------+---------+  31860
                    GATGGTCTAGCTCCACGAGCCGGAGTGTGGCTCGATACGCTACGCCGACTAGCTCTGCCG
      3              Y  Q  I  E  V  L  G  L  T  P  S  Y  A  M  R  L  I  E  T  A   -

                    CCGCGAGATGGGGCATCGACCTCGCCCGCGAGGCTAACGTCCAGATCATCCTGGCCGGGGC
      31861         ---------+---------+---------+---------+---------+---------+  31920
```

40

```
                 GGCGCTCTACCCGTAGCTGGAGCGGGCGCTCCGATTGCAGGTCTAGTAGGACCGGCCCCG
3                  R   E   M   G   I   D   L   A   R   E   A   N   V   Q   I   I   L   A   G   A  -

                 GGAGCCGCGCTCCGCGTTCACCACCCGCACCATCGAGGAGGCCTTCGGCGCCCGGGTCTT
31921            ---------+---------+---------+---------+---------+---------+ 31980
                 CCTCGGCGCGAGGCGCAAGTGGTGGGCGTGGTAGCTCCTCCGGAAGCCGCGGGCCCAGAA
3                  E   P   R   S   A   F   T   T   R   T   I   E   E   A   F   G   A   R   V   F  -

                 CAACGCCGCGGGCACCACTGAGTTCGGGGGGGTGTTCATGTTCGAGTGCACCGCCCGGCG
31981            ---------+---------+---------+---------+---------+---------+ 32040
                 GTTGCGGCGCCCGTGGTGACTCAAGCCCCCCCACAAGTACAAGCTCACGTGGCGGGCCGC
3                  N   A   A   G   T   T   E   F   G   G   V   F   M   F   E   C   T   A   R   R  -

                 CGAGGCCTGCCACATCATCGAACCCTCGTGCATCGAGGAGGTGCTCGACCCGGTGACGGA
32041            ---------+---------+---------+---------+---------+---------+ 32100
                 GCTCCGGACGGTGTAGTAGCTTGGGAGCACGTAGCTCCTCCACGAGCTGGGCCACTGCCT
3                  E   A   C   H   I   I   E   P   S   C   I   E   E   V   L   D   P   V   T   E  -

                 ACAGCCCGTCGGCTACGGCGAGGAGGGCGTCCGAGTCACCACCGGGCTGAACCGTGAGGG
32101            ---------+---------+---------+---------+---------+---------+ 32160
                 TGTCGGGCAGCCGATGCCGCTCCTCCCGCAGGCTCAGTGGTGGCCCGACTTGGCACTCCC
3                  Q   P   V   G   Y   G   E   E   G   V   R   V   T   T   G   L   N   R   E   G  -

                 GATGCAGCTCTTCCGGCACTGGACCGAGGACGTCGTGGTCAAGCGGCCCCACACCGAGTG
32161            ---------+---------+---------+---------+---------+---------+ 32220
                 CTACGTCGAGAAGGCCGTGACCTGGCTCCTGCAGCACCAGTTCGCCGGGGTGTGGCTCAC
3                  M   Q   L , F   R   H   W   T   E   D   V   V   V   K   R   P   H   T   E   C  -

                 CGGCTGCGGCCGGACGTGGGACTTCTACGACGGCGGCATCCTTCGGCGCGTGGACGACAT
32221            ---------+---------+---------+---------+---------+---------+ 32280
                 GCCGACGCCGGCCTGCACCCTGAAGATGCTGCCGCCGTAGGAAGCCGCGCACCTGCTGTA
3                  G   C   G   R   T   W   D   F   Y   D   G   G   I   L   R   R   V   D   D   M  -

                 GCGCAAGATACGCGGGGTCTCGATCACCCCGGTGATGATCGAGGATGTGCTGCGCGGCTT
32281            ---------+---------+---------+---------+---------+---------+ 32340
                 CGCGTTCTATGCGCCCCAGAGCTAGTGGGGCCACTACTAGCTCCTACACGACGCGCCGAA
3                  R   K   I   R   G   V   S   I   T   P   V   M   I   E   D   V   L   R   G   F  -

                 CGACGAGGTGAACGAGTTCCACTCGTCCATCCGGACCGTCCGCGGACTCGATACGATCCA
32341            ---------+---------+---------+---------+---------+---------+ 32400
                 GCTGCTCCACTTGCTCAAGGTGAGCAGGTAGGCCTGGCAGGCGCCTGAGCTATGCTAGGT
3                  D   E   V   N   E   F   H   S   S   I   R   T   V   R   G   L   D   T   I   H  -

                 CGTCAAGGTCGAGGCGGGAGACATCTCGGGTGAGGCGGCCGAGAGCCTGTGCGGCCGCAT
32401            ---------+---------+---------+---------+---------+---------+ 32460
                 GCAGTTCCAGCTCCGCCCTCTGTAGAGCCCACTCCGCCGGCTCTCGGACACGCCGGCGTA
3                  V   K   V   E   A   G   D   I   S   G   E   A   A   E   S   L   C   G   R   I  -

                 CACCGAGGAGTTCAAGCGTGAGATAGGCATACGGCCCCAGGTGGAGCTGACCCCCGCGGG
32461            ---------+---------+---------+---------+---------+---------+ 32520
                 GTGGCTCCTCAAGTTCGCACTCTATCCGTATGCCGGGGTCCACCTCGACTGGGGGCGCCC
3                  T   E   E   F   K   R   E   I   G   I   R   P   Q   V   E   L   T   P   A   G  -

                 CAGCCTCCCCCGATCGAAGTGGAAGGCGGCACGACTTCATGACGAGCGCGAACTCGCCCC
32521            ---------+---------+---------+---------+---------+---------+ 32580
                 GTCGGAGGGGGCTAGCTTCACCTTCCGCCGTGCTGAAGTACTGCTCGCGCTTGAGCGGGG
3                  S   L   P   R   S   K   W   K   A   A   R   L   H   D   E   R   E   L   A   P  -

                 TCAGGCCTGAGCAGGTGGAGCAGCTCCTGGTGAGCTACCGGAGCCTGGGCCTGCTGGAGC
32581            ---------+---------+---------+---------+---------+---------+ 32640
                 AGTCCGGACTCGTCCACCTCGTCGAGGACCACTCGATGGCCTCGGACCCGGACGACCTCG
3-*                Q   A   *  -

                 AGAGCTGCGCGGTCCCGGCCGTGCTCGCCGCGGTCAGGGCCGCCCGTGCGGAACTCCGTA
32641            ---------+---------+---------+---------+---------+---------+ 32700
                 TCTCGACGCGCCAGGGCCGGCACGAGCGGCGCCAGTCCCGGCGGGCACGCCTTGAGGCAT
```

41

```
                TCGCCCTGGACGGCCAGGGCGTGGAGTTCGAGTACTACCGGGGGCACGACGACAGCCTCG
         32701  ---------+---------+---------+---------+---------+---------+  32760
                AGCGGGACCTGCCGGTCCCGCACCTCAAGCTCATGATGGCCCCCGTGCTGCTGTCGGAGC

                TGGCCTGAACCCACCCCCGGTCCGCCGGGTCAGACGAAAGGGAGACCGGTGCCCCACGGT
         32761  ---------+---------+---------+---------+---------+---------+  32820
                ACCGGACTTGGGTGGGGGCCAGGCGGCCCAGTCTGCTTTCCCTCTGGCCACGGGGTGCCA
         4->                                                          M   P   H   G   -

                GCAGAGCGCGAAGCGAGCCCGGCCGAGGAGAGCGCCGGCACCCGGCCGCTGACCGGCGAG
         32821  ---------+---------+---------+---------+---------+---------+  32880
                CGTCTCGCGCTTCGCTCGGGCCGGCTCCTCTCGCGGCCGTGGGCCGGCGACTGGCCGCTC
         4      A   E   R   E   A   S   P   A   E   E   S   A   G   T   R   P   L   T   G   E   -

                GAGTATCTGGAGAGCCTGCGGGACGCGCGGGAGGTGTACCTCGACGGCAGCCGCGTCAAG
         32881  ---------+---------+---------+---------+---------+---------+  32940
                CTCATAGACCTCTCGGACGCCCTGCGCGCCCTCCACATGGAGCTGCCGTCGGCGCAGTTC
         4      E   Y   L   E   S   L   R   D   A   R   E   V   Y   L   D   G   S   R   V   K   -

                GACGTCACCGCGCATCCCGCGTTCCACAACCCGGCCCGGATGACGGCCCGGCTGTACGAC
         32941  ---------+---------+---------+---------+---------+---------+  33000
                CTGCAGTGGCGCGTAGGGCGCAAGGTGTTGGGCCGGGCCTACTGCCGGGCCGACATGCTG
         4      D   V   T   A   H   P   A   F   H   N   P   A   R   M   T   A   R   L   Y   D   -

                AGCCTGCACGACCCCGCCCAGAAAGCGGTCCTGACGGCGCCCACCGATGCCGGTGACGGT
         33001  ---------+---------+---------+---------+---------+---------+  33060
                TCGGACGTGCTGGGGCGGGTCTTTCGCCAGGACTGCCGCGGGTGGCTACGGCCACTGCCA
         4      S   L   H   D   P   A   Q   K   A   V   L   T   A   P   T   D   A   G   D   G   -

                TTCACCCACCGCTTCTTCACCGCACCGCGCAGCGTCGACGACCTGGTCAAGGACCAGGCC
         33061  ---------+---------+---------+---------+---------+---------+  33120
                AAGTGGGTGGCGAAGAAGTGGCGTGGCGCGTCGCAGCTGCTGGACCAGTTCCTGGTCCGG
         4      F   T   H   R   F   F   T   A   P   R   S   V   D   D   L   V   K   D   Q   A   -

                GCCATCGCATCCTGGGCGCGCAAGAGCTACGGCTGGATGGGGCGCAGCCCCGACTACAAG
         33121  ---------+---------+---------+---------+---------+---------+  33180
                CGGTAGCGTAGGACCCGCGCGTTCTCGATGCCGACCTACCCCGCGTCGGGGCTGATGTTC
         4      A   I   A   S   W   A   R   K   S   Y   G   W   M   G   R   S   P   D   Y   K   -

                GCGTCGTTCCTCGGCACGCTGGGGGCCAACGCCGACTTCTACGAGCCCTTCGCGGACAAC
         33181  ---------+---------+---------+---------+---------+---------+  33240
                CGCAGCAAGGAGCCGTGCGACCCCCGGTTGCGGCTGAAGATGCTCGGGAAGCGCCTGTTG
         4      A   S   F   L   G   T   L   G   A   N   A   D   F   Y   E   P   F   A   D   N   -

                GCCCGGCGCTGGTACCGGGAGTCGCAGGAGAAGGTGCTGTACTGGAACCATGCCTTCCTT
         33241  ---------+---------+---------+---------+---------+---------+  33300
                CGGGCCGCGACCATGGCCCTCAGCGTCCTCTTCCACGACATGACCTTGGTACGGAAGGAA
         4      A   R   R   W   Y   R   E   S   Q   E   K   V   L   Y   W   N   H   A   F   L   -

                CACCCGCCGGTCGACCGCTCGCTGCCCGCCGACGAGGTGGGCGACGTCTTCATCCACGTC
         33301  ---------+---------+---------+---------+---------+---------+  33360
                GTGGGCGGCCAGCTGGCGAGCGACGGGCGGCTGCTCCACCCGCTGCAGAAGTAGGTGCAG
         4      H   P   P   V   D   R   S   L   P   A   D   E   V   G   D   V   F   I   H   V   -

                GAGCGGGAGACCGACGCGGGCCTGGTGGTGAGCGGGGCCAAGGTCGTCGCGACCGGATCG
         33361  ---------+---------+---------+---------+---------+---------+  33420
                CTCGCCCTCTGGCTGCGCCCGGACCACCACTCGCCCCGGTTCCAGCAGCGCTGGCCTAGC
         4      E   R   E   T   D   A   G   L   V   V   S   G   A   K   V   V   A   T   G   S   -

                GCCCTCACCCACGCGGCGTTCATCTCGCACTGGGGACTTCCCATCAAGGACCGGAAGTTC
         33421  ---------+---------+---------+---------+---------+---------+  33480
                CGGGAGTGGGTGCGCCGCAAGTAGAGCGTGACCCCTGAAGGGTAGTTCCTGGCCTTCAAG
         4      A   L   T   H   A   A   F   I   S   H   W   G   L   P   I   K   D   R   K   F   -

                GCCCTGGTGGCCACCGTGCCGATGGACGCGGACGGCCTCAAGGTGATCTGCCGTCCCTCC
```

42

```
                                                                          ●                                      ●
       33481 ----------+---------+---------+---------+---------+---------+ 33540
             CGGGACCACCGGTGGCACGGCTACCTGCGCCTGCCGGAGTTCCACTAGACGGCAGGGAGG
       4     A  L  V  A  T  V  P  M  D  A  D  G  L  K  V  I  C  R  P  S    -

             TACTCCGCAAACGCGGCGACCACGGGCAGCCCGTTCGACAACCCGCTGTCCTCACGGCTG
       33541 ----------+---------+---------+---------+---------+---------+ 33600
             ATGAGGCGTTTGCGCCGCTGGTGCCCGTCGGGCAAGCTGTTGGGCGACAGGAGTGCCGAC
       4     Y  S  A  N  A  A  T  T  G  S  P  F  D  N  P  L  S  S  R  L    -

             GACGAGAACGACGCCATCCTCGTACTCGACCAGGTGCTGATCCCCTGGGAGAACGTGTTC
       33601 ----------+---------+---------+---------+---------+---------+ 33660
             CTGCTCTTGCTGCGGTAGGAGCATGAGCTGGTCCACGACTAGGGGACCCTCTTGCACAAG
       4     D  E  N  D  A  I  L  V  L  D  Q  V  L  I  P  W  E  N  V  F    -

             GTCTACGGCAACCTGGGCAAGGTACATCTCCTCGCCGGACAGTCCGGGATGATCGAACGC
       33661 ----------+---------+---------+---------+---------+---------+ 33720
             CAGATGCCGTTGGACCCGTTCCATGTAGAGGAGCGGCCTGTCAGGCCCTACTAGCTTGCG
       4     V  Y  G  N  L  G  K  V  H  L  L  A  G  Q  S  G  M  I  E  R    -

             GCCACCTTCCACGGGTGCACCCGGCTCGCCGTGAAGCTGGAGTTCATCGCCGGGCTGCTG
       33721 ----------+---------+---------+---------+---------+---------+ 33780
             CGGTGGAAGGTGCCCACGTGGGCCGAGCGGCACTTCGACCTCAAGTAGCGGCCCGACGAC
       4     A  T  F  H  G  C  T  R  L  A  V  K  L  E  F  I  A  G  L  L    -

             GCCAAGGCGCTGGACATCACCGGGGCGAAGGACTTCCGCGGTGTGCAGACCCGGCTCGGA
       33781 ----------+---------+---------+---------+---------+---------+ 33840
             CGGTTCCGCGACCTGTAGTGGCCCCGCTTCCTGAAGGCGCCACACGTCTGGGCCGAGCCT
       4     A  K  A  L  D  I  T  G  A  K  D  F  R  G  V  Q  T  R  L  G    -

             GAAGTCCTGGCCTGGCGCAACCTCTTCTGGTCACTGTCGGACGCGGCGGCCCGCAACCCC
       33841 ----------+---------+---------+---------+---------+---------+ 33900
             CTTCAGGACCGGACCGCGTTGGAGAAGACCAGTGACAGCCTGCGCCGCCGGGCGTTGGGG
       4     E  V  L  A  W  R  N  L  F  W  S  L  S  D  A  A  A  R  N  P    -

             GTCCCCTGGAAGAACGGCACGCTCCTGCCCAACCCTCAGGCGGGTATGGCCTACCGCTGG
       33901 ----------+---------+---------+---------+---------+---------+ 33960
             CAGGGGACCTTCTTGCCGTGCGAGGACGGGTTGGGAGTCCGCCCATACCGGATGGCGACC
       4     V  P  W  K  N  G  T  L  L  P  N  P  Q  A  G  M  A  Y  R  W    -

             TTCATGCAGATCGGCTACCCGCGGGTCCTGGAGATCGTCCAACAGGACGTGGCCAGCGGC
       33961 ----------+---------+---------+---------+---------+---------+ 34020
             AAGTACGTCTAGCCGATGGGCGCCCAGGACCTCTAGCAGGTTGTCCTGCACCGGTCGCCG
       4     F  M  Q  I  G  Y  P  R  V  L  E  I  V  Q  Q  D  V  A  S  G    -

             CTCATGTACGTCAACTCCTCCACGGAGGACTTCCGCAACCCCGAGACCGGCCCCTACTTG
       34021 ----------+---------+---------+---------+---------+---------+ 34080
             GAGTACATGCAGTTGAGGAGGTGCCTCCTGAAGGCGTTGGGGCTCTGGCCGGGGATGAAC
       4     L  M  Y  V  N  S  S  T  E  D  F  R  N  P  E  T  G  P  Y  L    -

             GAGAAGTACCTCCGGGGCAGCGACGGCGCAGGCGCCGTCGAGCGTGTCAAGGTGATGAAG
       34081 ----------+---------+---------+---------+---------+---------+ 34140
             CTCTTCATGGAGGCCCCGTCGCTGCCGCGTCCGCGGCAGCTCGCACAGTTCCACTACTTC
       4     E  K  Y  L  R  G  S  D  G  A  G  A  V  E  R  V  K  V  M  K    -

             CTGCTGTGGGACGCGGTGGGATCCGACTTCGGCGGCCGGCACGAACTCTACGAGCGGAAC
       34141 ----------+---------+---------+---------+---------+---------+ 34200
             GACGACACCCTGCGCCACCCTAGGCTGAAGCCGCCGGCCGTGCTTGAGATGCTCGCCTTG
       4     L  L  W  D  A  V  G  S  D  F  G  G  R  H  E  L  Y  E  R  N    -

             TACTCCGGGAACCACGAGAACACCCGGATCGAGTTGCTGCTGTCGCAGACGGCGAGCGGC
       34201 ----------+---------+---------+---------+---------+---------+ 34260
             ATGAGGCCCTTGGTGCTCTTGTGGGCCTAGCTCAACGACGACAGCGTCTGCCGCTCGCCG
       4     Y  S  G  N  H  E  N  T  R  I  E  L  L  L  S  Q  T  A  S  G    -

             AAACTGGACTCGTACATGGACTTCGCCCAGGCATGCATGGACGAGTACGACCTGGACGGC
       34261 ----------+---------+---------+---------+---------+---------+ 34320
```

43

```
              TTTGACCTGAGCATGTACCTGAAGCGGGTCCGTACGTACCTGCTCATGCTGGACCTGCCG
    4         K  L  D  S  Y  M  D  F  A  Q  A  C  M  D  E  Y  D  L  D  G  -

              TGGACCGCTCCCGACCTGGAGTCGTTTCACGCGATGCGTTCCGCCTCCCGCGACCTTCTC
    34321     ---------+---------+---------+---------+---------+---------+ 34380
              ACCTGGCGAGGGCTGGACCTCAGCAAAGTGCGCTACGCAAGGCGGAGGGCGCTGGAAGAG
    4         W  T  A  P  D  L  E  S  F  H  A  M  R  S  A  S  R  D  L  L  -

              GGAGGGCTGTAGTTCCCCGACGGTGTACTGCGGCCCCCGATCCGGGGGCCGCAGTACACC
    34381     ---------+---------+---------+---------+---------+---------+ 34440
              CCTCCCGACATCAAGGGGCTGCCACATGACGCCGGGGGCTAGGCCCCCGGCGTCATGTGG
    4-*       G  G  L  *  -

              GTCGGGGCGGCTGGTGCTCAGCCGCGCAGGAATCCGATGAGCTCGGGGGCGAGCTTCTTG
    34441     ---------+---------+---------+---------+---------+---------+ 34500
              CAGCCCCGCCGACCACGAGTCGGCGCGTCCTTAGGCTACTCGAGCCCCCGCTCGAAGAAC
    22-*            *  G  R  L  F  G  I  L  E  P  A  L  K  K  -

              GGCGCCATGGCGACGGCACCGTGGTTGAGCCCGTTCAGGGTGCGGTGGCTCGCGTCGGGG
    34501     ---------+---------+---------+---------+---------+---------+ 34560
              CCGCGGTACCGCTGCCGTGGCACCAACTCGGGCAAGTCCCACGCCACCGAGCGCAGCCCC
    22        P  A  M  A  V  A  G  H  N  L  G  N  L  T  R  H  S  A  D  P  -

              AGGACTCCGGTGAGTTCCTTCGCGGCACGCTGGAAACCGTCGGGGCTCTTGGAACCGGTC
    34561     ---------+---------+---------+---------+---------+---------+ 34620
              TCCTGAGGCCACTCAAGGAAGCGCCGTGCGACCTTTGGCAGCCCCGAGAACCTTGGCCAG
    22        L  V  G  T  L  E  K  A  A  R  Q  F  G  D  P  S  K  S  G  T  -

              AGCACCAGGGTCGGGGCCGACGCCGCCGACCACGGCTCGGCGGGGAGCGGCTTGCCCTGC
    34621     ---------+---------+---------+---------+---------+---------+ 34680
              TCGTGGTCCCAGCCCCGGCTGCGGCGGCTGGTGCCGAGCCGCCCCTCGCCGAACGGGACG
    22        L  V  L  T  P  A  S  A  A  S  W  P  E  A  P  L  P  K  G  Q  -

              TGGGTGTCGCCCATCACCGCGATGTCGTAGGGAAGCGTGTTGGCCAGACCCTTGAGGTTG
    34681     ---------+---------+---------+---------+---------+---------+ 34740
              ACCCACAGCGGGTAGTGGCGCTACAGCATCCCTTCGCACAACCGGTCTGGGAACTCCAAC
    22        Q  T  D  G  M  V  A  I  D  Y  P  L  T  N  A  L  G  K  L  N  -

              GACCAGACACCGGGCATCAGGCGCATGGCGCCGACCATGAAGGAGGGCATGCCCTGTGCC
    34741     ---------+---------+---------+---------+---------+---------+ 34800
              CTGGTCTGTGGCCCGTAGTCCGCGTACCGCGGCTGGTACTTCCTCCCGTACGGGACACGG
    22        S  W  V  G  P  M  L  R  M  A  G  V  M  F  S  P  M  G  Q  A  -

              TTGACCATGAAGGCCTTGACCGCGTCGCTGCGTCGGTCCTCCGCCAGAAGGCTGTCGATC
    34801     ---------+---------+---------+---------+---------+---------+ 34860
              AACTGGTACTTCCGGAACTGGCGCAGCGACGCAGCCAGGAGGCGGTCTTCCGACAGCTAG
    22        K  V  M  F  A  K  V  A  D  S  R  R  D  E  A  L  L  S  D  I  -

              TGACCGCCGAAGCCGGCGGGCGGGCCGAAGCCGTCCGAGGTGACGGAGAACGGCGGCTCG
    34861     ---------+---------+---------+---------+---------+---------+ 34920
              ACTGGCGGCTTCGGCCGCCCGCCCGGCTTCGGCAGGCTCCACTGCCTCTTGCCGCCGAGC
    22        Q  G  G  F  G  A  P  P  G  F  G  D  S  T  V  S  F  P  P  E  -

              TAGACCGCGAGCTTGTTCACCTTCAGGCCGGCGGCGGCGGCTCGCAGGGCGAGCACCGCG
    34921     ---------+---------+---------+---------+---------+---------+ 34980
              ATCTGGCGCTCGAACAAGTGGAAGTCCGGCCGCCGCCGCCGAGCGTCCCGCTCGTGGCGC
    22        Y  V  A  L  K  N  V  K  L  G  A  A  A  A  R  L  A  L  V  A  -

              CCGGAAGAGCTGCCGAACAGGGAGGCCGAACCGCCGACCTGGTCGATCAGCGCCGCGATG
    34981     ---------+---------+---------+---------+---------+---------+ 35040
              GGCCTTCTCGACGGCTTGTCCCTCCGGCTTGGCGGCTGGACCAGCTAGTCGCGGCGCTAC
    22        G  S  S  S  G  F  L  S  A  S  G  G  V  Q  D  I  L  A  A  I  -

              TCCTCGATCTCGCGCTCGACCGCGTACGCCGGACCGTCGGCGCTGGCGCCGCGGCCCCGA
    35041     ---------+---------+---------+---------+---------+---------+ 35100
              AGGAGCTAGAGCGCGAGCTGGCGCATGCGGCCTGGCAGCCGCGACCGCGGCGCCGGGGCT
```

44

```
22          D  E  I  E  R  E  V  A  Y  A  P. G  D  A  S  A  G  R  G  R  -
            CGGTCGTAGTTGACGACCGTGAAGTGCTCGGCGAGGAGACCGGCGAGCTTCTTGGCGTCG
     35101  ---------+---------+---------+---------+---------+---------+ 35160
            GCCAGCATCAACTGCTGGCACTTCACGAGCCGCTCCTCTGGCCGCTCGAAGAACCGCAGC
22          R  D  Y  N  V  V  T  F  H  E  A  L  L  G  A  L  K  K  A  D  -

            GAGCGGTCGGCCAAGGCGGAGGCCACCAGGATCACCGCCGGCCCCTCGCCCGACTTGTCG
     35161  ---------+---------+---------+---------+---------+---------+ 35220
            CTCGCCAGCCGGTTCCGCCTCCGGTGGTCCTAGTGGCGGCCGGGGAGCGGGCTGAACAGC
22          S  R  D  A  L  A  S  A  V  L  I  V  A  P  G  E  G  S  K  D  -

            AAGGCGATCGTGGTGCCGTCGGCCGATACCGTCGTTGATTCCACCTTGGCTGCTTTCTCA
     35221  ---------+---------+---------+---------+---------+---------+ 35280
            TTCCGCTAGCACCACGGCAGCCGGCTATGGCAGCAACTAAGGTGGAACCGACGAAAGAGT
22          F  I  A  T  T  G  D  A  S  V  T  T  S  E  V  K  A  A  K  E  -

            CGGGTTGAAGACATAGCTTCCCTCAGATCACATTGTGGGGCGTGCTGCCGACAGTGGAGA
     35281  ---------+---------+---------+---------+---------+---------+ 35340
            GCCCAACTTCTGTATCGAAGGGAGTCTAGTGTAACACCCCGCACGACGGCTGTCACCTCT
22-<         R  T  S  S  M  -

            CCGGCGTCCGGAGGAAAAGTAATCGGTCCTGCCAGAATTGGGGGTTCCGGAGGGCACGCC
     35341  ---------+---------+---------+---------+---------+---------+ 35400
            GGCCGCAGGCCTCCTTTTCATTAGCCAGGACGGTCTTAACCCCCAAGGCCTCCCGTGCGG

            GACCGCTGCACGACGGCGCGCCCCGACCTTCCGGACATTGTCGTGCCCTCAGATGTGTTT
     35401  ---------+---------+---------+---------+---------+---------+ 35460
            CTGGCGACGTGCTGCCGCGCGGGGCTGGAAGGCCTGTAACAGCACGGGAGTCTACACAAA

            CGCATCTTCAGGAGTGCTCAGTGATCCGTGAGGTGAGAAAGGGACGGTGGTCCGGTCAGT
     35461  ---------+---------+---------+---------+---------+---------+ 35520
            GCGTAGAAGTCCTCACGAGTCACTAGGCACTCCACTCTTTCCCTGCCACCAGGCCAGTCA
18-*                                                               *    -

            CGTTGCCGCGCGGGCTGTTCTGGTAAGCGGCCAGACGCCACTGCCCGTCCTGTTCGACGG
     35521  ---------+---------+---------+---------+---------+---------+ 35580
            GCAACGGCGCGCCCGACAAGACCATTCGCCGGTCTGCGGTGACGGGCAGGACAAGCTGCC
18           D  N  G  R  P  S  N  Q  Y  A  A  L  R  W  Q  G  D  Q  E  V  -

            CCAGCCAGGAGGCCCGGACGGCGCCGTCGCCGCTCGCCTCGGTCTCCCCCGGGGCGAGGA
     35581  ---------+---------+---------+---------+---------+---------+ 35640
            GGTCGGTCCTCCGGGCCTGCCGCGGCAGCGGCGAGCGGAGCCAGAGGGGGCCCCGCTCCT
18           A  L  W  S  A  R  V  A  G  D  G  S  A  E  T  E  G  P  A  L  -

            TGCCGCCCTCGGTGATGAGCAGGGCGATGCCGTCGCCGAGCAGGCGCGCGTCGATGGGGC
     35641  ---------+---------+---------+---------+---------+---------+ 35700
            ACGGCGGGAGCCACTACTCGTCCCGCTACGGCAGCGGCTCGTCCGCGCGCAGCTACCCCG
18           I  G  G  E  T  I  L  L  A  I  G  D  G  L  L  R  A  D  I  P  -

            TGCCGATGACACGGGTGCCCTTGTACGGGCCCGCGAAGGCGGCCGCCATGTGGGTGCGGA
     35701  ---------+---------+---------+---------+---------+---------+ 35760
            ACGGCTACTGTGCCCACGGGAACATGCCCGGGCGCTTCCGCCGGCGGTACACCCACGCCT
18           S  G  I  V  R  T  G  K  Y  P  G  A  F  A  A  A  M  H  T  R  -

            TGTTCTCGCGGCCCTTGCGGAAGAGGCCGGGGAGGATCATCGTCCCGTCCTCGGCGAAGA
     35761  ---------+---------+---------+---------+---------+---------+ 35820
            ACAAGAGCGCCGGGAACGCCTTCTCCGGCCCCTCCTAGTAGCAGGGCAGGAGCCGCTTCT
18           I  N  E  R  G  K  R  F  L  G  P  L  I  M  T  G  D  E  A  F  -

            CGTCGGCGAACCGGTCGGCGTCGTGGTCGGCCCAGGCGGCCACGATGCGCGCCGGCAGAG
     35821  ---------+---------+---------+---------+---------+---------+ 35880
            GCAGCCGCTTGGCCAGCCGCAGCACCAGCCGGGTCCGCCGGTGCTACGCGCGGCCGTCTC
18           V  D  A  F  R  D  A  D  H  D  A  W  A  A  V  I  R  A  P  L  -

            CGGCTACCGCTGCCAGGGCGGCGTCGGGAGCGGAGGTGGTCGAGTCGGTGCTGGTCATAT
```

45

```
                    ●                              ●

35881  ---------+---------+---------+---------+---------+---------+  35940
       GCCGATGGCGACGGTCCCGCCGCAGCCCTCGCCTCCACCAGCTCAGCCACGACCAGTATA
  18-<    A  A  V  A  A  L  A  A  D  P  A  S  T  T  S  D  T  S  T  M  -

       CGCGGTTCCCGTCCGTTGGTTGGCGGTTTCGGCACGGCCCGCAGCCCTGCCCGAGCCCGA
35941  ---------+---------+---------+---------+---------+---------+  36000
       GCGCCAAGGGCAGGCAACCAACCGCCAAAGCCGTGCCGGGCGTCGGGACGGGCTCGGGCT

       CGCTGGCAGGCGGCCCCGTCATCAGGCATCTCCTGCGTTGCGCCCCACGCCAGTCACTTC
36001  ---------+---------+---------+---------+---------+---------+  36060
       GCGACCGTCCGCCGGGGCAGTAGTCCGTAGAGGACGCAACGCGGGGTGCGGTCAGTGAAG

       ACGGCCAGAACAAGTCGCGCATTCTGGAAGAAGCTGAGGCCCGCGACCCGGTGCGACGAT
36061  ---------+---------+---------+---------+---------+---------+  36120
       TGCCGGTCTTGTTCAGCGCGTAAGACCTTCTTCGACTCCGGGCGCTGGGCCACGCTGCTA

       CTGCGGTGTCACGGAGTTCGCACACGTTTACGCACGGAGGCTCGATGCCCGCTGTCAATG
36121  ---------+---------+---------+---------+---------+---------+  36180
       GACGCCACAGTGCCTCAAGCGTGTGCAAATGCGTGCCTCCGAGCTACGGGCGACAGTTAC
   5->                                                M  P  A  V  N  G  -

       GATCGGTGCAGTCAGGCCAGTCGCACCGACGCTCCGTCGTGGCGACGGTGGTGGGCAACT
36181  ---------+---------+---------+---------+---------+---------+  36240
       CTAGCCACGTCAGTCCGGTCAGCGTGGCTGCGAGGCAGCACCGCTGCCACCACCCGTTGA
   5      S  V  Q  S  G  Q  S  H  R  R  S  V  V  A  T  V  V  G  N  F  -

       TCGTGGAGTCGTTCGACTGGCTCGCCTACGGGCTCTTCGCTCCTCTCTTCGCGGCTCAGT
36241  ---------+---------+---------+---------+---------+---------+  36300
       AGCACCTCAGCAAGCTGACCGAGCGGATGCCCGAGAAGCGAGGAGAGAAGCGCCGAGTCA
   5      V  E  S  F  D  W  L  A  Y  G  L  F  A  P  L  F  A  A  Q  F  -

       TCTTCCCCTCGTCCAACCAGTTCACCTCCCTGCTCGGCGCGTTCGCGGTCTTCGGCACGG
36301  ---------+---------+---------+---------+---------+---------+  36360
       AGAAGGGGAGCAGGTTGGTCAAGTGGAGGGACGAGCCGCGCAAGCGCCAGAAGCCGTGCC
   5      F  P  S  S  N  Q  F  T  S  L  L  G  A  F  A  V  F  G  T  G  -

       GCATGCTCTTCCGGCCGATCGGCGGGGTCCTGCTGGGCCGCCTCGCCGACCGGCGCGGCC
36361  ---------+---------+---------+---------+---------+---------+  36420
       CGTACGAGAAGGCCGGCTAGCCGCCCCAGGACGACCCGGCGGAGCGGCTGGCCGCGCCGG
   5      M  L  F  R  P  I  G  G  V  L  L  G  R  L  A  D  R  R  G  R  -

       GGCGCCCCGCCCTGATGCTGGCGATCGGACTGATGACCGGCGGCTCGACCCTGATCGCCG
36421  ---------+---------+---------+---------+---------+---------+  36480
       CCGCGGGGCGGGACTACGACCGCTAGCCTGACTACTGGCCGCCGAGCTGGGACTAGCGGC
   5      R  P  A  L  M  L  A  I  G  L  M  T  G  G  S  T  L  I  A  V  -

       TCGTCCCCACCTACGAGCACATCGGGATCCTCGCCCCGCTGCTTCTGCTGCTCGCCCGGC
36481  ---------+---------+---------+---------+---------+---------+  36540
       AGCAGGGGTGGATGCTCGTGTAGCCCTAGGAGCGGGGCGACGAAGACGACGAGCGGGCCG
   5      V  P  T  Y  E  H  I  G  I  L  A  P  L  L  L  L  L  A  R  L  -

       TCGCCCAGGGAGTCTCCTCGGGCGGGGAATGGACAGCGGCGGCCACCTACCTGATGGAGA
36541  ---------+---------+---------+---------+---------+---------+  36600
       AGCGGGTCCCTCAGAGGAGCCCGCCCCTTACCTGTCGCCGCCGGTGGATGGACTACCTCT
   5      A  Q  G  V  S  S  G  G  E  W  T  A  A  A  T  Y  L  M  E  I  -

       TCGCGCCGAAGAACCGCCGGTGCCTCTACAGCAGCCTCTTCTCCGTGACGACCATGGCGG
36601  ---------+---------+---------+---------+---------+---------+  36660
       AGCGCGGCTTCTTGGCGGCCACGGAGATGTCGTCGGAGAAGAGGCACTGCTGGTACCGCC
   5      A  P  K  N  R  R  C  L  Y  S  S  L  F  S  V  T  T  M  A  G  -

       GCCCCTTCGTCGCATCGCTGCTGGGCGCGGGCCTCGGCGTGTGGCTGGGAACCGCGACGA
36661  ---------+---------+---------+---------+---------+---------+  36720
       CGGGGAAGCAGCGTAGCGACGACCCGCGCCCGGAGCCGCACACCGACCCTTGGCGCTGCT
   5      P  F  V  A  S  L  L  G  A  G  L  G  V  W  L  G  T  A  T  M  -
```

46

```
             TGGAGGCCTGGGGCTGGCGGGTGCCGTTCCTCCTCGGCGGCGTCTTCGGCGTGATCCTGC
      36721  ---------+---------+---------+---------+---------+---------+  36780
             ACCTCCGGACCCCGACCGCCCACGGCAAGGAGGAGCCGCCGCAGAAGCCGCACTAGGACG
      5         E  A  W  G  W  R  V  P  F  L  L  G  G  V  F  G  V  I  L  L  -

             TGTTCCTGCGCCGTCGGCTCACCGAGACCGAGGTCTTCCGCCGGGAGGTGCGGCCCCGGG
      36781  ---------+---------+---------+---------+---------+---------+  36840
             ACAAGGACGCGGCAGCCGAGTGGCTCTGGCTCCAGAAGGCGGCCCTCCACGCCGGGGCCC
      5         F  L  R  R  R  L  T  E  T  E  V  F  R  R  E  V  R  P  R  A  -

             CCCGGCGCGGCTCACTGGGCCAGCTGATCGGAGCCCACCGCCCCCAGGTGCTGCTGGCCG
      36841  ---------+---------+---------+---------+---------+---------+  36900
             GGGCCGCGCCGAGTGACCCGGTCGACTAGCCTCGGGTGGCGGGGGTCCACGACGACCGGC
      5         R  R  G  S  L  G  Q  L  I  G  A  H  R  P  Q  V  L  L  A  V  -

             TGATGCTGGTGGCCGGACTGGGCGTCATCGGCGGAACGTGGTCGACCGCGGTCCCGGCGA
      36901  ---------+---------+---------+---------+---------+---------+  36960
             ACTACGACCACCGGCCTGACCCGCAGTAGCCGCCTTGCACCAGCTGGCGCCAGGGCCGCT
      5         M  L  V  A  G  L  G  V  I  G  G  T  W  S  T  A  V  P  A  M  -

             TGGGCCACCGTCTGATCGGCTCGCAGACGATGTTCTGGGTGGTGGTCTGTGTGACCGGCT
      36961  ---------+---------+---------+---------+---------+---------+  37020
             ACCCGGTGGCAGACTAGCCGAGCGTCTGCTACAAGACCCACCACCAGACACACTGGCCGA
      5         G  H  R  L  I  G  S  Q  T  M  F  W  V  V  V  C  V  T  G  S  -

             CGGTCATCCTGCTGCAGGTACCCATAGGGCTGCTCGCCGACCGGGTGGAACCGGGCAGGT
      37021  ---------+---------+---------+---------+---------+---------+  37080
             GCCAGTAGGACGACGTCCATGGGTATCCCGACGAGCGGCTGGCCCACCTTGGCCCGTCCA
      5         V  I  L  L  Q  V  P  I  G  L  L  A  D  R  V  E  P  G  R  F  -

             TCCTGATCGTCTCCAGCGTCGTCTTCGCCGCTGTGGGCTCGTACGCCTACCTCACCGTCC
      37081  ---------+---------+---------+---------+---------+---------+  37140
             AGGACTAGCAGAGGTCGCAGCAGAAGCGGCGACACCCGAGCATGCGGATGGAGTGGCAGG
      5         L  I  V  S  S  V  V  F  A  A  V  G  S  Y  A  Y  L  T  V  Q  -

             AGGACTCCTTCGCGAGCCTGGCGTTCACGTACAGCACCGGAGTGATCTTCCTCGGCTGCG
      37141  ---------+---------+---------+---------+---------+---------+  37200
             TCCTGAGGAAGCGCTCGGACCGCAAGTGCATGTCGTGGCCTCACTAGAAGGAGCCGACGC
      5         D  S  F  A  S  L  A  F  T  Y  S  T  G  V  I  F  L  G  C  V  -

             TCACCATGGTGCTGCCGAAGATGCTCTCCAGAATCTTCCCTCCGCAGATACGCGGCCTGG
      37201  ---------+---------+---------+---------+---------+---------+  37260
             AGTGGTACCACGACGGCTTCTACGAGAGGTCTTAGAAGGGAGGCGTCTATGCGCCGGACC
      5         T  M  V  L  P  K  M  L  S  R  I  F  P  P  Q  I  R  G  L  G  -

             GCATCGGGCTGCCGCACGCCTCGACCACCGCACTCCTCGGCGGGGCGGGGCCACTGCTGG
      37261  ---------+---------+---------+---------+---------+---------+  37320
             CGTAGCCCGACGGCGTGCGGAGCTGGTGGCGTGAGGAGCCGCCCCGCCCCGGTGACGACC
      5         I  G  L  P  H  A  S  T  T  A  L  L  G  G  A  G  P  L  L  A  -

             CCGCCTACTCCGACGAGCGAGGCGCCTCGGGCTGGTTCATCGCCGCCGTGATGGCCGCGG
      37321  ---------+---------+---------+---------+---------+---------+  37380
             GGCGGATGAGGCTGCTCGCTCCGCGGAGCCCGACCAAGTAGCGGCGGCACTACCGGCGCC
      5         A  Y  S  D  E  R  G  A  S  G  W  F  I  A  A  V  M  A  A  V  -

             TCCTGCTCGCCTGGCCGGCCACCCTGTGGGAGCGACGGCTGTTCCGCGCCCGGACGGCCC
      37381  ---------+---------+---------+---------+---------+---------+  37440
             AGGACGAGCGGACCGGCCGGTGGGACACCCTCGCTGCCGACAAGGCGCGGGCCTGCCGGG
      5         L  L  A  W  P  A  T  L  W  E  R  R  L  F  R  A  R  T  A  P  -

             CGGGAAGCGAGCCGGTTCCCGAATCCGCCGTCGCCCGCCCCGTCGGGTGACCGTCCGCAC
      37441  ---------+---------+---------+---------+---------+---------+  37500
             GCCCTTCGCTCGGCCAAGGGCTTAGGCGGCAGCGGGCGGGGCAGCCCACTGGCAGGCGTG
      5-*       G  S  E  P  V  P  E  S  A  V  A  R  P  V  G  *  -
```

47

```
              TTCTGCATCCCGTCCGGCACCGAGCGCCGGCGACCTTCCCGACTGAGAGGTTGACATCAT
    37501     ---------+---------+---------+---------+---------+---------+  37560
              AAGACGTAGGGCAGGCCGTGGCTCGCGGCCGCTGGAAGGGCTGACTCTCCAACTGTAGTA
    23->                                                                  M  -

              GACGACGTCCGACACCACCGACCGGTCCCAGGACGGCGTGCCGCCGCTCTCCTTCCACCA
    37561     ---------+---------+---------+---------+---------+---------+  37620
              CTGCTGCAGGCTGTGGTGGCTGGCCAGGGTCCTGCCGCACGGCGGCGAGAGGAAGGTGGT
    23          T  T  S  D  T  T  D  R  S  Q  D  G  V  P  P  L  S  F  H  Q  -

              GGAGTTCCTGTGCATGTTCGACAGCGGGAACGACGGCGCCGACGTGGGGCCGTTCGGCCC
    37621     ---------+---------+---------+---------+---------+---------+  37680
              CCTCAAGGACACGTACAAGCTGTCGCCCTTGCTGCCGCGGCTGCACCCCGGCAAGCCGGG
    23          E  F  L  C  M  F  D  S  G  N  D  G  A  D  V  G  P  F  G  P  -

              CATGTACCACATCGTCGGAGCCTGGCGGCTGACCGGCGGGATCGACGAGGAGACCCTGCG
    37681     ---------+---------+---------+---------+---------+---------+  37740
              GTACATGGTGTAGCAGCCTCGGACCGCCGACTGGCCGCCCTAGCTGCTCCTCTGGGACGC
    23          M  Y  H  I  V  G  A  W  R  L  T  G  G  I  D  E  E  T  L  R  -

              CGAGGCGCTGGGTGACGTCGTCGTGCGCCACGAGGCCCTGCGCACATCGCTGGTCCGCGA
    37741     ---------+---------+---------+---------+---------+---------+  37800
              GCTCCGCGACCCACTGCAGCAGCACGCGGTGCTCCGGGACGCGTGTAGCGACCAGGCGCT
    23          E  A  L  G  D  V  V  V  R  H  E  A  L  R  T  S  L  V  R  E  -

              AGGTGGCACGCACCGGCCGGAGATCCTGCCTGCGGGGCCCGCCGCGCTGGAGGTCCGTGA
    37801     ---------+---------+---------+---------+---------+---------+  37860
              TCCACCGTGCGTGGCCGGCCTCTAGGACGGACGCCCCGGGCGGCGCGACCTCCAGGCACT
    23          G  G  T  H  R  P  E  I  L  P  A  G  P  A  A  L  E  V  R  D  -

              TCTCGGCGACGTCGACGAGTCGGAGCGGGTGCGGCGCGGTGAGGAACTGCTCAACGAGGT
    37861     ---------+---------+---------+---------+---------+---------+  37920
              AGAGCCGCTGCAGCTGCTCAGCCTCGCCCACGCCGCGCCACTCCTTGACGAGTTGCTCCA
    23          L  G  D  V  D  E  S  E  R  V  R  R  G  E  E  L  L  N  E  V  -

              GGAGTCGACCGGTCTGAGCGTGCGGGAGCTGCCCCTGCTGCGGGCCGTGCTCGGACGCTT
    37921     ---------+---------+---------+---------+---------+---------+  37980
              CCTCAGCTGGCCAGACTCGCACGCCCTCGACGGGGACGACGCCCGGCACGAGCCTGCGAA
    23          E  S  T  G  L  S  V  R  E  L  P  L  L  R  A  V  L  G  R  F  -

              CGACCAGAAGGACGCGGTGCTGGTCCTCATCGCCCACCACACCGCCGCGGACGCCTGGGC
    37981     ---------+---------+---------+---------+---------+---------+  38040
              GCTGGTCTTCCTGCGCCACGACCAGGAGTAGCGGGTGGTGTGGCGGCGCCTGCGGACCCG
    23          D  Q  K  D  A  V  L  V  L  I  A  H  H  T  A  A  D  A  W  A  -

              CATGCACGTCATCGCCCGCGACCTGCTCAACCTGTACGCCGCCAGGCGCGGGAACCCGGT
    38041     ---------+---------+---------+---------+---------+---------+  38100
              GTACGTGCAGTAGCGGGCGCTGGACGAGTTGGACATGCGGCGGTCCGCGCCCTTGGGCCA
    23          M  H  V  I  A  R  D  L  L  N  L  Y  A  A  R  R  G  N  P  V  -

              TCCCCCGCTCCCCGAGCCGGCCCAGCATGCCGAGTTCGCCCGCTGGGAGCGCGAGGCGGC
    38101     ---------+---------+---------+---------+---------+---------+  38160
              AGGGGGCGAGGGGCTCGGCCGGGTCGTACGGCTCAAGCGGGCGACCCTCGCGCTCCGCCG
    23          P  P  L  P  E  P  A  Q  H  A  E  F  A  R  W  E  R  E  A  A  -

              CGAGGCACCGCGGGTCGCGGTCTCGAAGGAATTCTGGCGCAAGCGCCTCCAGGGCGCGCG
    38161     ---------+---------+---------+---------+---------+---------+  38220
              GCTCCGTGGCGCCCAGCGCCAGAGCTTCCTTAAGACCGCGTTCGCGGAGGTCCCGCGCGC
    23          E  A  P  R  V  A  V  S  K  E  F  W  R  K  R  L  Q  G  A  R  -

              GATCATCGGGCTGGAGACGGACATACCGCGCTCGGCGGGGCTGCCCAAGGGCACCGCGTG
    38221     ---------+---------+---------+---------+---------+---------+  38280
              CTAGTAGCCCGACCTCTGCCTGTATGGCGCGAGCCGCCCCGACGGGTTCCCGTGGCGCAC
    23          I  I  G  L  E  T  D  I  P  R  S  A  G  L  P  K  G  T  A  W  -

              GCAGCGCTTCGCCGTACGCGGGGAACTGGCCGACGCCGTGGTGGAGTTCTCACGGGCCGC
```

```
                     ●                               ●


         38281 ---------+---------+---------+---------+---------+---------+ 38340
               CGTCGCGAAGCGGCATGCGCCCCTTGACCGGCTGCGGCACCACCTCAAGAGTGCCCGGCG
           23   Q  R  F  A  V  R  G  E  L  A  D  A  V  V  E  F  S  R  A  A  -

               CAAGTGCTCCCCGTTCATGACCATGTTCGCCGCCTACCAGGTGCTGCTGCACCGCAGGAC
         38341 ---------+---------+---------+---------+---------+---------+ 38400
               GTTCACGAGGGGCAAGTACTGGTACAAGCGGCGGATGGTCCACGACGACGTGGCGTCCTG
           23   K  C  S  P  F  M  T  M  F  A  A  Y  Q  V  L  L  H  R  R  T  -

               GGGCGAGCTGGACATCACCGTGCCGACCTTCTCCGGGGGGCGCAACAACTCGCGGTTCGA
         38401 ---------+---------+---------+---------+---------+---------+ 38460
               CCCGCTCGACCTGTAGTGGCACGGCTGGAAGAGGCCCCCCGCGTTGTTGAGCGCCAAGCT
           23   G  E  L  D  I  T  V  P  T  F  S  G  G  R  N  N  S  R  F  E  -

               GGACACCGTCGGTTCCTTCATCAACTTCCTGCCGCTGCGTACCGACCTCTCCGGATGCGC
         38461 ---------+---------+---------+---------+---------+---------+ 38520
               CCTGTGGCAGCCAAGGAAGTAGTTGAAGGACGGCGACGCATGGCTGGAGAGGCCTACGCG
           23   D  T  V  G  S  F  I  N  F  L  P  L  R  T  D  L  S  G  C  A  -

               ATCCTTCCGCGAGGTCGTGCTGCGCACCCGCACCACCTGCGGAGAGGCGTTCACCCACGA
         38521 ---------+---------+---------+---------+---------+---------+ 38580
               TAGGAAGGCGCTCCAGCACGACGCGTGGGCGTGGTGGACGCCTCTCCGCAAGTGGGTGCT
           23   S  F  R  E  V  V  L  R  T  R  T  T  C  G  E  A  F  T  H  E  -

               GCTGCCCTTCTCCCGGCTGATCCCGGAGGTGCCGGAGCTGATGGCGTCGGCGGCCTCCGA
         38581 ---------+---------+---------+---------+---------+---------+ 38640
               CGACGGGAAGAGGGCCGACTAGGGCCTCCACGGCCTCGACTACCGCAGCCGCCGGAGGCT
           23   L  P  F  S  R  L  I  P  E  V  P  E  L  M  A  S  A  A  S  D  -

               CAACCACCAGATCTCCGTCTTCCAGGCCGTGCACGCGCCCGCGTCCGAGGGGCCCGAGCA
         38641 ---------+---------+---------+---------+---------+---------+ 38700
               GTTGGTGGTCTAGAGGCAGAAGGTCCGGCACGTGCGCGGGCGCAGGCTCCCCGGGCTCGT
           23   N  H  Q  I  S  V  F  Q  A  V  H  A  P  A  S  E  G  P  E  Q  -

               GGCCGGGGACCTGACGTACTCGAAGATCTGGGAGCGGCAGCTGTCGCAGGCGGAGGGCTC
         38701 ---------+---------+---------+---------+---------+---------+ 38760
               CCGGCCCCTGGACTGCATGAGCTTCTAGACCCTCGCCGTCGACAGCGTCCGCCTCCCGAG
           23   A  G  D  L  T  Y  S  K  I  W  E  R  Q  L  S  Q  A  E  G  S  -

               CGACATCCCCGACGGGGTGCTGTGGTCGATCCACATCGACCCCTCGGGCTCCATGGCCGG
         38761 ---------+---------+---------+---------+---------+---------+ 38820
               GCTGTAGGGGCTGCCCCACGACACCAGCTAGGTGTAGCTGGGGAGCCCGAGGTACCGGCC
           23   D  I  P  D  G  V  L  W  S  I  H  I  D  P  S  G  S  M  A  G  -

               CAGCCTCGGGTACAACACCAACCGCTTCAAGGACGAGACGATGGCGGCCTTCCTGGCCGA
         38821 ---------+---------+---------+---------+---------+---------+ 38880
               GTCGGAGCCCATGTTGTGGTTGGCGAAGTTCCTGCTCTGCTACCGCCGGAAGGACCGGCT
           23   S  L  G  Y  N  T  N  R  F  K  D  E  T  M  A  A  F  L  A  D  -

               CTACCTCGACGTGCTCGAGAACGCGGTGGCCCGGCCGGACGCCCCCTTCACCTCCTGAGA
         38881 ---------+---------+---------+---------+---------+---------+ 38940
               GATGGAGCTGCACGAGCTCTTGCGCCACCGGGCCGGCCTGCGGGGGAAGTGGAGGACTCT
         23-*   Y  L  D  V  L  E  N  A  V  A  R  P  D  A  P  F  T  S  *  -

               CAGTTCCGGCGGCGGCGAACCCGCCCGAAGAAAGGAAAGCCAGTGTCCACCGTTTCCGAC
         38941 ---------+---------+---------+---------+---------+---------+ 39000
               GTCAAGGCCGCCGCCGCTTGGGCGGGCTTCTTTCCTTTCGGTCACAGGTGGCAAAGGCTG
         26->                                                    M  S  T  V  S  D  -

               ACAGCGGCCGGCTCCTCCCTGGAGGAGAAGGTCACCCGGATCTGGACGGGTGTTCTCGGC
         39001 ---------+---------+---------+---------+---------+---------+ 39060
               TGTCGCCGGCCGAGGAGGGACCTCCTCTTCCAGTGGGCCTAGACCTGCCCACAAGAGCCG
           26   T  A  A  G  S  S  L  E  E  K  V  T  R  I  W  T  G  V  L  G  -

               ACGTCCGGTGAGGAAGGCGCGACGTTCATCGAGCTCGGAGGGCAGTCGGTCTCGGCCGTG
         39061 ---------+---------+---------+---------+---------+---------+ 39120
```

49

```
                     TGCAGGCCACTCCTTCCGCGCTGCAAGTAGCTCGAGCCTCCCGTCAGCCAGAGCCGGCAC
26                     T  S  G  E  E  G  A  T  F  I  E  L  G  G  Q  S  V  S  A  V   -

                     CGCATCGCCACGCGTATCCAGGAGGAGCTCGACATCTGGGTCGACATCGGCGTCCTCTTC
39121     ---------+---------+---------+---------+---------+---------+  39180
                     GCGTAGCGGTGCGCATAGGTCCTCCTCGAGCTGTAGACCCAGCTGTAGCCGCAGGAGAAG
26                     R  I  A  T  R  I  Q  E  E  L  D  I  W  V  D  I  G  V  L  F   -

                     GACGACCCGGATCTGCCTACCTTCATCGCGGCGGTCGTCCGGACGGCCGACGCCGCGGGC
39181     ---------+---------+---------+---------+---------+---------+  39240
                     CTGCTGGGCCTAGACGGATGGAAGTAGCGCCGCCAGCAGGCCTGCCGGCTGCGGCGCCCG
26                     D  D  P  D  L  P  T  F  I  A  A  V  V  R  T  A  D  A  A  G   -

                     GGCGAGGGCTCCGGAACGCAGTGAGACTCGCCGGGCGCCGTCTCCCCGCGGCGCCCGGTT
39241     ---------+---------+---------+---------+---------+---------+  39300
                     CCGCTCCCGAGGCCTTGCGTCACTCTGAGCGGCCCGCGGCAGAGGGGCGCCGCGGGCCAA
26-*                   G  E  G  S  G  T  Q  *  -

                     TCACATGGCTGAGGCGGTTCACCCGGTACCGGGTGAACCGCCTCAGCCATGTGAAACCGG
39301     ---------+---------+---------+---------+---------+---------+  39360
                     AGTGTACCGACTCCGCCAAGTGGGCCATGGCCCACTTGGCGGAGTCGGTACACTTTGGCC

                     GCCTGGTCAGCGCAGCTGGATGTCCGTCTCCCGGGCGATCGCCCGGAGGAACTCGCCGCG
39361     ---------+---------+---------+---------+---------+---------+  39420
                     CGGACCAGTCGCGTCGACCTACAGGCAGAGGGCCCGCTAGCGGGCCTCCTTGAGCGGCGC
24-*                      *  R  L  Q  I  D  T  E  R  A  I  A  R  L  F  E  G  R   -

                     GGACAGCGCGTCGGCGACCAGCTCGATGTCGTCGGCCATGTACCGGTCGACGCCCAGCGT
39421     ---------+---------+---------+---------+---------+---------+  39480
                     CCTGTCGCGCAGCCGCTGGTCGAGCTACAGCAGCCGGTACATGGCCAGCTGCGGGTCGCA
24                     S  L  A  D  A  V  L  E  I  D  D  A  M  Y  R  D  V  G  L  T  -

                     CGGAACCAGCCGGCGCACCGCTTCGTACGTGGCCTTCGCCGCCGGGCTCAAGCCGTCGAA
39481     ---------+---------+---------+---------+---------+---------+  39540
                     GCCTTGGTCGGCCGCGTGGCGAAGCATGCACCGGAAGCGGCGGCCCGAGTTCGGCAGCTT
24                     P  V  L  R  R  V  A  E  Y  T  A  K  A  A  P  S  L  G  D  F  -

                     CCGGCCGGAGATGTCGACCGCCTGGGCGGCGGCCAGGTACTCCACCGCGAGGATCTTGTT
39541     ---------+---------+---------+---------+---------+---------+  39600
                     GGCCGGCCTCTACAGCTGGCGGACCCGCCGCCGGTCCATGAGGTGGCGCTCCTAGAACAA
24                     R  G  S  I  D  V  A  Q  A  A  A  L  Y  E  V  A  L  I  K  N  -

                     GTTGTTCGACAGGACCCGGCGGGCGTTGCGGGCCGAGATCAGGCCCATGCTCACCACGTC
39601     ---------+---------+---------+---------+---------+---------+  39660
                     CAACAAGCTGTCCTGGGCCGCCCGCAACGCCCGGCTCTAGTCCGGGTACGAGTGGTGCAG
24                     N  N  S  L  V  R  R  A  N  R  A  S  I  L  G  M  S  V  V  D  -

                     CTGGTTGTCGCCGTTGGACGGGACGCTCTGGGTGCTGGCCGGGCCGATCGTCCGGTTCTC
39661     ---------+---------+---------+---------+---------+---------+  39720
                     GACCAACAGCGGCAACCTGCCCTGCGAGACCCACGACCGGCCCGGCTAGCAGGCCAAGAG
24                     Q  N  D  G  N  S  P  V  S  Q  T  S  A  P  G  I  T  R  N  E  -

                     GGCCACCAGTGCGGTGGCCGGGTACTGGGCGCCGGCGAATCCGCTGTGCAGCCCCGGGTC
39721     ---------+---------+---------+---------+---------+---------+  39780
                     CCGGTGGTCACGCCACCGGCCCATGACCCGCGGCCGCTTAGGCGACACGTCGGGGCCCAG
24                     A  V  L  A  T  A  P  Y  Q  A  G  A  F  G  S  H  L  G  P  D  -

                     CCCGGAGACGAGGAACTCCGGGAGGCCGTAGCTGAGGTGCCGGTTCAGGACCCGGTTGAT
39781     ---------+---------+---------+---------+---------+---------+  39840
                     GGGCCTCTGCTCCTTGAGGCCCTCCGGCATCGACTCCACGGCCAAGTCCTGGGCCAACTA
24                     G  S  V  L  F  E  P  L  G  Y  S  L  H  R  N  L  V  R  N  I  -

                     CTGCCGCTCGGCCAGGACGCCGAGCTGGGTGAGCGCGATGGTCACGAAGTCCATCGCGAA
39841     ---------+---------+---------+---------+---------+---------+  39900
                     GACGGCGAGCCGGTCCTGCGGCTCGACCCACTCGCGCTACCAGTGCTTCAGGTAGCGCTT
```

```
24          Q  R  E  A  L  V  G  L  Q  T  L  A  I  T  V  F  D  M  A  F -

            CGCGATCGGCTGACCGTGGAAGTTCGCCCCGTGGAAGATCTCCTTGCCCTCGAAGAAGAG
   39901    ---------+---------+---------+---------+---------+---------+ 39960
            GCGCTAGCCGACTGGCACCTTCAAGCGGGGCACCTTCTAGAGGAACGGGAGCTTCTTCTC
24           A  I  P  Q  G  H  F  N  A  G  H  F  I  E  K  G  E  F  F  L -

            CGGGTTGTCGTTGGCCGAGTTGAGCTCGATGCGCAGCTTGTGCCGCGCGTGGTACAAGGT
   39961    ---------+---------+---------+---------+---------+---------+ 40020
            GCCCAACAGCAACCGGCTCAACTCGAGCTACGCGTCGAACACGGCGCGCACCATGTTCCA
24           P  N  D  N  A  S  N  L  E  I  R  L  K  H  R  A  H  Y  L  T -

            GTCGCGCACCGCCCCGACGACCTGGGGGATGGCCCGCAGCGAGTAGGCCTTCTGCAGGTA
   40021    ---------+---------+---------+---------+---------+---------+ 40080
            CAGCGCGTGGCGGGGCTGCTGGACCCCCTACCGGGCGTCGCTCATCCGGAAGACGTCCAT
24            D  R  V  A  G  V  V  Q  P  I  A  R  L  S  Y  A  K  Q  L  Y -

            GATCTCCGAGCGCTGGACGTCCTTGCCGGCCTCCTTGTCCTTCTGGAGTTCTCGGCGCAG
   40081    ---------+---------+---------+---------+---------+---------+ 40140
            CTAGAGGCTCGCGACCTGCAGGAACGGCCGGAGGAACAGGAAGACCTCAAGAGCCGCGTC
24            I  E  S  R  Q  V  D  K  G  A  E  K  D  K  Q  L  E  R  R  L -

            GTCGGCGTGCTCGACCGTCAGTCCGCTGCCCCGCATCAGGGCCCGCATGTTGGCGGCGGT
   40141    ---------+---------+---------+---------+---------+---------+ 40200
            CAGCCGCACGAGCTGGCAGTCAGGCGACGGGGCGTAGTCCCGGGCGTACAACCGCCGCCA
24            D  A  H  E  V  T  L  G  S  G  R  M  L  A  R  M  N  A  A  T -

            GTCGATCTGGCCCTCGTGCGGGCGGGCTATGTCGTGCCCCTCCGCGAGGAAGGGGCTGGT
   40201    ---------+---------+---------+---------+---------+---------+ 40260
            CAGCTAGACCGGGAGCACGCCCGCCCGATACAGCACGGGGAGGCGCTCCTTCCCCGACCA
24            D  I  Q  G  E  H  P  R  A  I  D  H  G  E  A  L  F  P  S  T -

            CGATCCGCGTACCGCCTCGATGAGCAGAGCCGTCACGATCTCGGCCTGCTGGGCCTGCTC
   40261    ---------+---------+---------+---------+---------+---------+ 40320
            GCTAGGCGCATGGCGGAGCTACTCGTCTCGGCAGTGCTAGAGCCGGACGACCCGGACGAG
24            S  G  R  V  A  E  I  L  L  A  T  V  I  E  A  Q  Q  A  Q  E -

            CAGGGCCCGTCCGACGACCAGGGAGCCCAGACCGGTCATCCCGGACGTGCCGTTGATCAG
   40321    ---------+---------+---------+---------+---------+---------+ 40380
            GTCCCGGGCAGGCTGCTGGTCCCTCGGGTCTGGCCAGTAGGGCCTGCACGGCAACTAGTC
24            L  A  R  G  V  V  L  S  G  L  G  T  M  G  S  T  G  N  I  L -

            TGCGAGGCCCTCCTTGAAGCGCAGTTCGAGCGGCTCGATGCCCCGCTCGGCCAGCACCTG
   40381    ---------+---------+---------+---------+---------+---------+ 40440
            ACGCTCCGGGAGGAACTTCGCGTCAAGCTCGCCGAGCTACGGGGCGAGCCGGTCGTGGAC
24            A  L  G  E  K  F  R  L  E  L  P  E  I  G  R  E  A  L  V  Q -

            GGCGGTCTCCACCGGCCGTCCGTCGCGCAGGACGTAGCCCTCTCCGATGAGGGTGCTCGC
   40441    ---------+---------+---------+---------+---------+---------+ 40500
            CCGCCAGAGGTGGCCGGCAGGCAGCGCGTCCTGCATCGGGAGAGGCTACTCCCACGAGCG
24            A  T  E  V  P  R  G  D  R  L  V  Y  G  E  G  I  L  T  S  A -

            GACGTGGGAGAGGGGAGCCAGGTCGCCGCTCGCCCCGAGTGACCCGATCTCGGGTATGGC
   40501    ---------+---------+---------+---------+---------+---------+ 40560
            CTGCACCCTCTCCCCTCGGTCCAGCGGCGAGCGGGGCTCACTGGGCTAGAGCCCATACCG
24            V  H  S  L  P  A  L  D  G  S  A  G  L  S  G  I  E  P  I  A -

            CGGGGTGATGCCCTCGTTCAGGTACTGCGCGAGGCGTTCGAGGATGATGGGGCGCACCGC
   40561    ---------+---------+---------+---------+---------+---------+ 40620
            GCCCCACTACGGGAGCAAGTCCATGACGCGCTCCGCAAGCTCCTACTACCCCGCGTGGCG

24            P  T  I  G  E  N  L  Y  Q  A  L  R  E  L  I  I  P  R  V  A -

            GGAGTGGCCCTTGGCGAGGGTGTTCAGCCGGGCGGCGACGATCGCCCGCGCCTCGTCCTC
   40621    ---------+---------+---------+---------+---------+---------+ 40680
            CCTCACCGGGAACCGCTCCCACAAGTCGGCCCGCCGCTGCTAGCGGGCGCGGAGCAGGAG
```

51

24   S H G K A L T N L R A A V I A R A E D E-

```
       GGCGAACAGCGGACCGACTCCCGCGCTGTGGCTACGGACGAGATTGGTCTGCAGTTCGAC
40681  ---------+---------+---------+---------+---------+---------+ 40740
       CCGCTTGTCGCCTGGCTGAGGGCGCGACACCGATGCCTGCTCTAACCAGACGTCAAGCTG
24        A   F   L   P   G   V   G   A   S   H   S   R   V   L   N   T   Q   L   E   V-

       TTCCTTCGACTTGTCGACCTGCATGTAGATCATCTCGCCGTACCCGGTGGTCACCCCGTA
40741  ---------+---------+---------+---------+---------+---------+ 40800
       AAGGAAGCTGAACAGCTGGACGTACATCTAGTAGAGCGGCATGGGCCACCAGTGGGGCAT
24        E   K   S   K   D   V   Q   M   Y   I   M   E   G   Y   G   T   T   V   G   Y-

       GATGGGGATGTTCTGTTCGGCGATCCCTTCGAAGATCTCCCGGCTCTTCTGGGCCTTCGC
40801  ---------+---------+---------+---------+---------+---------+ 40860
       CTACCCCTACAAGACAAGCCGCTAGGGAAGCTTCTAGAGGGCCGAGAAGACCCGGAAGCG
24        I   P   I   N   Q   E   A   I   G   E   F   I   E   R   S   K   Q   A   K   A-

       GATGGATTCGGCCGGTACGTCGACCGTCGCGCGTTCCTCCGCGACGCGGCGTACGGCTTC
40861  ---------+---------+---------+---------+---------+---------+ 40920
       CTACCTAAGCCGGCCATGCAGCTGGCAGCGCGCAAGGAGGCGCTGCGCCGCATGCCGAAG
24        I   S   E   A   P   V   D   V   T   A   R   E   E   A   V   R   R   V   A   E-

       GACGGTCAGGGTCTCGCCGTCGACGGAAACCGGGACGATCTCGGTCTCGACTTGAGTCAA
40921  ---------+---------+---------+---------+---------+---------+ 40980
       CTGCCAGTCCCAGAGCGGCAGCTGCCTTTGGCCCTGCTAGAGCCAGAGCTGAACTCAGTT
24         V   T   L   T   E   G   D   V   S   V   P   V   I   E   T   E   V   Q   T   L-

       TGCCATCACTCCATGGGTAGCGGCCGAGGCCGGTGTACGACAGGTCAGGGGGTGGGTTCG
40981  ---------+---------+---------+---------+---------+---------+ 41040
       ACGGTAGTGAGGTACCCATCGCCGGCTCCGGCCACATGCTGTCCAGTCCCCCACCCAAGC
24-<       A   M-

       TGAGGCGCGGCTCAGCGGGTGAGCCGGGAGCGGTCCACCTTCCCCGCGGCGTTGCGCGGC
41041  ---------+---------+---------+---------+---------+---------+ 41100
       ACTCCGCGCCGAGTCGCCCACTCGGCCCTCGCCAGGTGGAAGGGGCGCCGCAACGCGCCG
25-*            *   R   T   L   R   S   R   D   V   K   G   A   A   N   R   P-

       AGGCGTGAAGTCAGGCGGGTGAAGACGGCGGGCAGTGCGAGGGGGCCGAACTGGCCGCGC
41101  ---------+---------+---------+---------+---------+---------+ 41160
       TCCGCACTTCAGTCCGCCCACTTCTGCCGCCCGTCACGCTCCCCCGGCTTGACCGGCGCG
25        L   R   S   T   L   R   T   F   V   A   P   L   A   L   P   G   F   Q   G   R-

       AGATGGGAACGCCAGGCCCGGATGTCCGCGCGCACGTCCTCCCGGCCCTCTCCTTGTGGC
41161  ---------+---------+---------+---------+---------+---------+ 41220
       TCTACCCTTGCGGTCCGGGCCTACAGGCGCGCGTGCAGGAGGGCCGGGAGAGGAACACCG
25        L   H   S   R   W   A   R   I   D   A   R   V   D   E   R   G   E   G   Q   P-

       ACCACGTACACGGCGAGGCGGGTCACCAGGCCCTGGCCGTTGACGTGGGGGAGGACCGCG
41221  ---------+---------+---------+---------+---------+---------+ 41280
       TGGTGCATGTGCCGCTCCGCCCAGTGGTCCGGGACCGGCAACTGCACCCCCTCCTGGCGC
25        V   V   Y   V   A   L   R   T   V   L   G   Q   G   N   V   H   P   L   V   A-

       CACTCCAGGACCGAGGGGTCACGGTTCAGCGCGGCCTCGATCTCGGTGAGTTCCAAGCGG
41281  ---------+---------+---------+---------+---------+---------+ 41340
       GTGAGGTCCTGGCTCCCCAGTGCCAAGTCGCGCCGGAGCTAGAGCCACTCAAGGTTCGCC
25        C   E   L   V   S   P   D   R   N   L   A   A   E   I   E   T   L   E   L   R-

       TTCCCGAACAGCTTGACCTGGAAGTCCTTGCGGCCCCGGAATTCCAGGGCTCCGTCGAAC
41341  ---------+---------+---------+---------+---------+---------+ 41400
       AAGGGCTTGTCGAACTGGACCTTCAGGAACGCCGGGGCCTTAAGGTCCCGAGGCAGCTTG
25        N   G   F   L   K   V   Q   F   D   K   R   G   R   F   E   L   A   G   D   F-

       CGTACCCGCGCCAGATCCCCGGTCCGGTACCACCGGTCACCGTCCGGGGCGAGGCCGGCG
41401  ---------+---------+---------+---------+---------+---------+ 41460
       GCATGGGCGCGGTCTAGGGGCCAGGCCATGGTGGCCAGTGGCAGGCCCCGCTCCGGCCGC
25        R   V   R   A   L   D   G   T   R   Y   W   R   D   G   D   P   A   L   G   A-
```

52

```
            AGGGGCGCGAACAGCGCGCTGTGGTCCGGGCCGCCCTCGACGGCGAGATAACCCGGCGTC
    41461   ---------+---------+---------+---------+---------+---------+ 41520
            TCCCCGCGCTTGTCGCGCGACACCAGGCCCGGCGGGAGCTGCCGCTCTATTGGGCCGCAG
    25       L  P  A  F  L  A  S  H  D  P  G  G  E  V  A  L  Y  G  P  T  -

            ACGTACGGGGAGCGGATCACCAGTTCGCCGGTGACGCCGGCGGGGCTCGGCCGGTCGTCC
    41521   ---------+---------+---------+---------+---------+---------+ 41580
            TGCATGCCCCTCGCCTAGTGGTCAAGCGGCCACTGCGGCCGCCCCGAGCCGGCCAGCAGG
    25       V  Y  P  S  R  I  V  L  E  G  T  V  G  A  P  S  R  D  D  -

            GCGTCCACGACGAGTACCTGGCGGCCGGGGAGCGGGTACCCGATCGGGGCCGGGCCCGTG
    41581   ---------+---------+---------+---------+---------+---------+ 41640
            CGCAGGTGCTGCTCATGGACCGCCGGCCCCTCGCCCATGGGCTAGCCCCGGCCCGGGCAC
    25       A  D  V  V  L  V  Q  R  G  P  L  P  Y  G  I  P  A  P  G  T  -

            ACCGGCCCGGTGATCTCGTGCCAGGTCGCGGCGATCGTCTCGGTGGGCCCGTAGAGGTTG
    41641   ---------+---------+---------+---------+---------+---------+ 41700
            TGGCCGGGCCACTAGAGCACGGTCCAGCGCCGCTAGCAGAGCCACCCGGGCATCTCCAAC
    25       V  P  G  T  I  E  H  W  T  A  A  I  T  E  T  P  G  Y  L  N  -

            ATCAGGCGGGTCCGGGGCAGGGCCGCGCGCAGTCCGTCCACGAGTTCGCCGGGCAGCGCC
    41701   ---------+---------+---------+---------+---------+---------+ 41760
            TAGTCCGCCCAGGCCCCGTCCCGGCGCGCGTCAGGCAGGTGCTCAAGCGGCCCGTCGCGG
    25       I  L  R  T  R  P  L  A  A  R  L  G  D  V  L  E  G  P  L  A  -

            TCGCCCATCAGGAGCAGGTGGCCCAGGGTGCCGGGCCGATCGCCCGGGTCGGAGGCGGTG
    41761   ---------+---------+---------+---------+---------+---------+ 41820
            AGCGGGTAGTCCTCGTCCACCGGGTCCCACGGCCCGGCTAGCGGGCCCAGCCTCCGCCAC
    25       E  G  M  L  L  L  H  G  L  T  G  P  R  D  G  P  D  S  A  T  -

            ATCACTCCCAGGAGGTCCCGGGCGAAGCTGGGCACGGTCTGGAGATGAGTGATCCGCTCC
    41821   ---------+---------+---------+---------+---------+---------+ 41880
            TAGTGAGGGTCCTCCAGGGCCCGCTTCGACCCGTGCCAGACCTCTACTCACTAGGCGAGG
    25       I  V  G  L  L  D  R  A  F  S  P  V  T  Q  L  H  T  I  R  E  -

            TGGACGAGCCACGGCACCAGCTTGTCGGGGTTCACCCTGACGCGCTCCGGCACCGGACAC
    41881   ---------+---------+---------+---------+---------+---------+ 41940
            ACCTGCTCGGTGCCGTGGTCGAACAGCCCCAAGTGGGACTGCGCGAGGCCGTGGCCTGTG
    25       Q  V  L  W  P  V  L  K  D  P  N  V  R  V  R  E  P  V  P  C  -

            AGCGTCCCGCCGGCCACGAGCGTCGCGAAGACCTCGGCCAGCGCCGGGTCGTGCTCCGGG
    41941   ---------+---------+---------+---------+---------+---------+ 42000
            TCGCAGGGCGGCCGGTGCTCGCAGCGCTTCTGGAGCCGGTCGCGGCCCAGCACGAGGCCC
    25       L  T  G  G  A  V  L  T  A  F  V  E  A  L  A  P  D  H  E  P  -
```

SEQ ID No. 2. C-1027 gene cluster DNA sequence from 41,980 to 63,164

```
            AGCGCCGGGTCGTGCTCCGGGGAGACCCACTGCGCCACCCGCGCGCCCGGCCCCATCGCG
    41980   +---------+---------+---------+---------+--------+---------- 42039
            TCGCGGCCCAGCACGAGGCCCCTCTGGGTGACGCGGTGGGCGCGCGGGCCGGGGTAGCGC
    25       L  A  P  D  H  E  P  S  V  W  Q  A  V  R  A  G  P  G  M  A  -

            AACCGTTCGCCCATCCAGCCCGCGAACTGGCCCAGCGCGGCATGCGACTGGGCGATCCCC
    42040   +---------+---------+---------+---------+--------+---------- 42099
            TTGGCAAGCGGGTAGGTCGGGCGCTTGACCGGGTCGCGCCGTACGCTGACCCGCTAGGGG
    25       F  R  E  G  M  W  G  A  F  Q  G  L  A  A  H  S  Q  A  I  G  -

            TTGGGCCGCCCGGTCGAACCCGAGGTGAACGCCACGTAGGCCAGGTCTGCCAGGCCCGGC
    42100   +---------+---------+---------+---------+--------+---------- 42159
            AACCCGGCGGGCCAGCTTGGGCTCCACTTGCGGTGCATCCGGTCCAGACGGTCCGGGCCG
```

●   ●

```
25              K  P  R  G  T  S  G  S  T  F  A  V  Y  A  L  D  A  L  G  P  -

                CCCGCCGCGGTCGTCGCGTCCGGGCCGGCGGCGGGTCGAGGGCCGAGCACAGAGGAGGCG
        42160   +--------+---------+---------+---------+---------+---------   42219
                GGGCGGCGCCAGCAGCGCAGGCCCGGCCGCCGCCCAGCTCCCGGCTCGTGTCCTCCGC
        25        G  A  A  T  T  A  D  P  G  A  A  P  R  P  G  L  V  S  S  A  -

                TCCAGCAGGGTGGCGCCCGGTTCACCGGCGTACCAGAGCGCCAGCGGATCCTCCTGCGGA
        42220   +--------+---------+---------+---------+---------+---------   42279
                AGGTCGTCCCACCGCGGGCCAAGTGGCCGCATGGTCTCGCGGTCGCCTAGGAGGACGCCT
        25         D  L  L  T  A  G  P  E  G  A  Y  W  L  A  L  P  D  E  Q  P  -

                TCGCCGTCGAGGACCAGGCACGCCGGGCGCAGATCGCTGAGCATCGACCGGTGTCGTTCG
        42280   +--------+---------+---------+---------+---------+---------   42339
                AGCGGCAGCTCCTGGTCCGTGCGGCCCGCGTCTAGCGACTCGTAGCTGGCCACAGCAAGC
        25         D  G  D  L  V  L  C  A  P  R  L  D  S  L  M  S  R  H  R  E  -

                CCCGCGCCGTCCGGAGCGAACCACGCCAGGTGGGCGCCCGCCTCCAGGACTCCCAGCAGC
        42340   +--------+---------+---------+---------+---------+---------   42399
                GGGCGCGGCAGGCCTCGCTTGGTGCGGTCCACCCGCGGGCGGAGGTCCTGAGGGTCGTCG
        25        G  A  G  D  P  A  F  W  A  L  H  A  G  A  E  L  V  G  L  L  -

                ACCGCGATCCGGCGGGCGCCCGGCTGCATCCGCACCGCCACCGGCGAGCCGTGCCCCGCG
        42400   +--------+---------+---------+---------+---------+---------   42459
                TGGCGCTAGGCCGCCCGCGGGCCGACGTAGGCGTGGCGGTGGCCGCTCGGCACGGGGCGC
        25         V  A  I  R  R  A  G  P  Q  M  R  V  A  V  P  S  G  H  G  A  -

                CCGGCCGCGGTGAGGGCCGAGGCGACGCGGGCCGCGTCCGCGGTCAGTTCGGCGGTCAGT
        42460   +--------+---------+---------+---------+---------+---------   42519
                GGCCGGCGCCACTCCCGGCTCCGCTGCGCCCGGCGCAGGCGCCAGTCAAGGCGCCAGTCA
        25         G  A  A  T  L  A  S  A  V  R  A  A  D  A  T  L  E  A  T  L  -

                TCGGCGTAGCTTGTGCGCGTGCCGCCGAACGAGACGGCGACACCGTCGTGTTCCGCGTGG
        42520   +--------+---------+---------+---------+---------+---------   42579
                AGCCGCATCGAACACGCGCACGGCGGCTTGCTCTGCCGCTGTGGCAGCACAAGGCGCACC
        25       E  A  Y  S  T  R  T  G  G  F  S  V  A  V  G  D  H  E  A  H  -

                CGGCGGACCGAGGCGTGCACCGGCCGCGTCATGTCCCCGCCGGACGCCCGGCGGTCCGAA
        42580   +--------+---------+---------+---------+---------+---------   42639
                GCCGCCTGGCTCCGCACGTGGCCGGCGCAGTACAGGGGCGGCCTGCGGGCCGCCAGGCTT
        25       R  R  V  S  A  H  V  P  R  T  M  (ORF25)

                GCGCGCAGGGCGTGGTCCCGGTGGCGGTCGTCGTCCAGCGGCAGAGCGCCCACGGGTGTG
        42640   +--------+---------+---------+---------+---------+---------   42699
                CGCGCGTCCCGCACCAGGGCCACCGCCAGCAGCAGGTCGCCGTCTCGCGGGTGCCCACAC

                TCCGGATCCGTGGTCGCGGCGGTCAGGAGGACGGCCAGCTGATCCAGCATCCGCCGGGCC
        42700   +--------+---------+---------+---------+---------+---------   42759
                AGGCCTAGGCACCAGCGCCGCCAGTCCTCCTGCCGGTCGACTAGGTCGTAGGCGGCCCGG

                GAAGCGGGCTCGAACAGAGCTTCGCGGTACTCCAGGTAGCCGGTGACCGAGGGCGCGGTG
        42760   +--------+---------+---------+---------+---------+---------   42819
                CTTCGCCCGAGCTTGTCTCGAAGCGCCATGAGGTCCATCGGCCACTGGCTCCCGCGCCAC

                TCCTGCAGCACCAGGGTCAGGTCGGCGGCGGCAGTGCCGTTGTGCACGGACAGCCGCCTC
        42820   +--------+---------+---------+---------+---------+---------   42879
                AGGACGTCGTGGTCCCAGTCCAGCCGCCGCCGTCACGGCAACACGTGCCTGTCGGCGGAG

                ACCTCGGCGCCTGGTATCCGCAGGCCCGGCCGCTCCTCGTGGACGAACACGGCGTCGGCC
        42880   +--------+---------+---------+---------+---------+---------   42939
                TGGAGCCGCGGACCATAGGCGTCCGGGCCGGCGAGGAGCACCTGCTTGTGCCGCAGCCGG

                CCCTCGATCCGGCACGGCCCGGGGGCCGGGGCCGGCGTCGTGTGCAGCAGCTCCCGGAAG
        42940   +--------+---------+---------+---------+---------+---------   42999
                GGGAGCTAGGCCGTGCCGGGCCCCCGGCCCCGGCCGCAGCACACGTCGTCGAGGGCCTTC
```

54

```
       GCGGTGGCCGGCGTGCCGTCGTCCTGTCCGGCGTAGCGCTGGACCAGGGCTCGGAATCCG
43000  +---------+---------+---------+---------+---------+---------  43059
       CGCCACCGGCCGCACGGCAGCAGGACAGGCCGCATCGCGACCTGGTCCCGAGCCTTAGGC

       GCCAGCACCACGGCCGCGGCGGTGACCCCTTCCGCTTCGGCGAGCCGGGCCGTACGGAAG
43060  +---------+---------+---------+---------+---------+---------  43119
       CGGTCGTGGTGCCGGCGCCGCCACTGGGGAAGGCGAAGCCGCTCGGCCCGGCATGCCTTC

       CCGAGGTCCGGACTCCAGCCGAAGGCGACGGTGCTCCCCGCGTGCGAGGGCAGGTGCGGG
43120  +---------+---------+---------+---------+---------+---------  43179
       GGCTCCAGGCCTGAGGTCGGCTTCCGCTGCCACGAGGGGCGCACGCTCCCGTCCACGCCC

       CGGTTCCGGTCGGCGGGCAGGACCTGTCCGGAGGCGGTCGCCGAAGACTCCTCGCTCCCG
43180  +---------+---------+---------+---------+---------+---------  43239
       GCCAAGGCCAGCCGCCCGTCCTGGACAGGCCTCCGCCAGCGGCTTCTGAGGAGCGAGGGC

       GGCGCCCGGGGCGTTTGCGGCGCGGGCGCAGTGGGAGGCCGGCCGCCGGTGGTGACGGCG
43240  +---------+---------+---------+---------+---------+---------  43299
       CCGCGGGCCCCGCAAACGCCGCGCCCGCGTCACCCTCCGGCCGGCGGCCACCACTGCCGC

       AGGTACGCGTTCGACAACGCGGCCGGCAGGGGCCCGGACGGCCCGTCCCAGGCTCCGGAG
43300  +---------+---------+---------+---------+---------+---------  43359
       TCCATGCGCAAGCTGTTGCGCCGGCCGTCCCCGGGCCTGCCGGGCAGGGTCCGAGGCCTC

       TGCGAGGCCACCAGGAGAAGCAGGTGCGCGCGTGGGCCTCTGCGGGCGATGTGGAGCCGT
43360  +---------+---------+---------+---------+---------+---------  43419
       ACGCTCCGGTGGTCCTCTTCGTCCACGCGCGCACCCGGAGACGCCCGCTACACCTCGGCA

       GCGGGCGCGTCACCCTCGGCGAAGGGACGGGCCGCCCAGCGAGCGCAGAGTTCCTCCTCC
43420  +---------+---------+---------+---------+---------+---------  43479
       CGCCCGCGCAGTGGGAGCCGCTTCCCTGCCCGGCGGGTCGCTCGCGTCTCAAGGAGGAGG

       CCGCACTCCTCGTCGGCACTCGGCCCGTCCACGGCGGCCCCGTCTCCGGCGGCGGCCCGC
43480  +---------+---------+---------+---------+---------+---------  43539
       GGCGTGAGGAGCAGCCGTGAGCCGGGCAGGTGCCGCCGGGGCAGAGGCCGCCGCCGGGCG

       CAGGCCGTCCGCAGGGCCTCCAGGTCGAGTCCGCCGCTCACGTGGTAGGCCGCGTACGGG
43540  +---------+---------+---------+---------+---------+---------  43599
       GTCCGGCAGGCGTCCCGGAGGTCCAGCTCAGGCGGCGAGTGCACCATCCGGCGCATGCCC

       TGCAACACCGCAGATCCGGAGGCCGGCGAAGGCCCCCGGTCCGGCTCGGTCACAGTCACG
43600  +---------+---------+---------+---------+---------+---------  43659
       ACGTTGTGGCGTCTAGGCCTCCGGCCGCTTCCGGGGGCCAGGCCGAGCCAGTGTCAGTGC

       TCATTCGCCACGACGCCCATCTTGGGGCGGCGGCGCACAGGACGCTTCTCCTTGAGTGCG
43660  +---------+---------+---------+---------+---------+---------  43719
       AGTAAGCGGTGCTGCGGGTAGAACCCCGCCGCCGCGTGTCCTGCGAAGAGGAACTCACGC

       GAGCTCCGCGTACGGCGCCGAAGCGTTCGGTCAAACCTTGTTCGACCAACTGCGCAATCT
43720  +---------+---------+---------+---------+---------+---------  43779
       CTCGAGGCGCATGCCGCGGCTTCGCAAGCCAGTTTGGAACAAGCTGGTTGACGCGTTAGA

       GGAAGTTGACGTCTTCCAGGTGGAGTTGGGAACGATGGAGGCCCCCGCCGGCCGCGTCGG
43780  +---------+---------+---------+---------+---------+---------  43839
       CCTTCAACTGCAGAAGGTCCACCTCAACCCTTGCTACCTCCGGGGGCGGCCGGCGCAGCC

       AACGGCCGTGCAGTGCGGCCCTCTCCAACACTCCCGGCCATCGCGGAATCCGAGACGTGC
43840  +---------+---------+---------+---------+---------+---------  43899
       TTGCCGGCACGTCACGCCGGGAGAGGTTGTGAGGGCCGGTAGCGCCTTAGGCTCTGCACG

       CCGAAGGAGCCCCCCTTGCAAGCCTGGTTCAAGCGCACCAGTGGTGTGCCCGGTGACAGA
43900  +---------+---------+---------+---------+---------+---------  43959
       GGCTTCCTCGGGGGGAACGTTCGGACCAAGTTCGCGTGGTCACCACACGGGCCACTGTCT
27                                                 (ORF27)  V  P  G  D  R
-
```

55

```
              CGTGGAAAGTGGCTGGTCCTGGCCGCCTGGCTCATCATCGCGATGGCGCTGGGCCCGCTG
      43960   +---------+---------+---------+---------+---------+---------  44019
              GCACCTTTCACCGACCAGGACCGGCGGACCGAGTAGTAGCGCTACCGCGACCCGGGCGAC
      27          R  G  K  W  L  V  L  A  A  W  L  I  I  A  M  A  L  G  P  L   -

              GCGGGGAAGCTCGCCGACGTCCAGGACTCCAGCGCCAACGCCTTCCTTCCGCGCAGCTCG
      44020   +---------+---------+---------+---------+---------+---------  44079
              CGCCCCTTCGAGCGGCTGCAGGTCCTGAGGTCGCGGTTGCGGAAGGAAGGCGCGTCGAGC
      27         A  G  K  L  A  D  V  Q  D  S  S  A  N  A  F  L  P  R  S  S   -

              GAGTCCGCGAAGCTGAACAAGGAACTGGAGAAGTTCCGCGCCGACGAGCTGATGCCGGCC
      44080   +---------+---------+---------+---------+---------+---------  44139
              CTCAGGCGCTTCGACTTGTTCCTTGACCTCTTCAAGGCGCGGCTGCTCGACTACGGCCGG
      27         E  S  A  K  L  N  K  E  L  E  K  F  R  A  D  E  L  M  P  A   -

              GTGGTGGTCTACAGCGCCGACGGCTCGCTGCCCGCCGAGGGGCGGGCCAAGGCCGAGAAG
      44140   +---------+---------+---------+---------+---------+---------  44199
              CACCACCAGATGTCGCGGCTGCCGAGCGACGGGCGGCTCCCCGCCCGGTTCCGGCTCTTC
      27         V  V  V  Y  S  A  D  G  S  L  P  A  E  G  R  A  K  A  E  K   -

              GACATAGCCGCCTTCCAGGAGCTGGCCGCCGAGGGCGAGAAGGTCGAAGCGCCCCTGGAG
      44200   +---------+---------+---------+---------+---------+---------  44259
              CTGTATCGGCGGAAGGTCCTCGACCGGCGGCTCCCGCTCTTCCAGCTTCGCGGGGACCTC
      27         D  I  A  A  F  Q  E  L  A  A  E  G  E  K  V  E  A  P  L  E   -

              TCGGAGGACGGCCAGGCGCTCATGGTCGTCGTTCCGCTGATCAGCGACGCCGACATCGTC
      44260   +---------+---------+---------+---------+---------+---------  44319
              AGCCTCCTGCCGGTCCGCGAGTACCAGCAGCAAGGCGACTAGTCGCTGCGGCTGTAGCAG
      27         S  E  D  G  Q  A  L  M  V  V  V  P  L  I  S  D  A  D  I  V   -

              GCCACGACGAAGAAGGTCCGCGATGTCGCGGACGCCAACGCCCCCCCGGGCGTCGCCATC
      44320   +---------+---------+---------+---------+---------+---------  44379
              CGGTGCTGCTTCTTCCAGGCGCTACAGCGCCTGCGGTTGCGGGGGGGCCCGCAGCGGTAG
      27         A  T  T  K  K  V  R  D  V  A  D  A  N  A  P  P  G  V  A  I   -

              GAGGTGGGCGGGCCCGCCGGGTCGACGACCGACGCCGCCGGCGCTTTCGAGTCCCTCGAC
      44380   +---------+---------+---------+---------+---------+---------  44439
              CTCCACCCGCCCGGGCGGCCCAGCTGCTGGCTGCGGCGGCCGCGAAAGCTCAGGGAGCTG
      27         E  V  G  G  P  A  G  S  T  T  D  A  A  G  A  F  E  S  L  D   -

              TCCATGCTGATGATGGTCACCGGCCTTGTGGTCGCCATCCTGCTGCTGATCACCTACCGC
      44440   +---------+---------+---------+---------+---------+---------  44499
              AGGTACGACTACTACCAGTGGCCGGAACACCAGCGGTAGGACGACGACTAGTGGATGGCG
      27         S  M  L  M  M  V  T  G  L  V  V  A  I  L  L  L  I  T  Y  R   -

              TCCCCCATCCTGTGGCTGCTGCCCCTGCTCTCCGTCGGCTTCGCCTCCGTGCTGACCCAG
      44500   +---------+---------+---------+---------+---------+---------  44559
              AGGGGGTAGGACACCGACGACGGGGACGAGAGGCAGCCGAAGCGGAGGCACGACTGGGTC
      27         S  P  I  L  W  L  L  P  L  L  S  V  G  F  A  S  V  L  T  Q   -

              GTCGGCACCTACATGCTCGCCAAGTACGCCGGGCTGCCGGTCGACCCGCAGAGCTCCGGC
      44560   +---------+---------+---------+---------+---------+---------  44619
              CAGCCGTGGATGTACGAGCGGTTCATGCGGCCCGACGGCCAGCTGGGCGTCTCGAGGCCG
      27         V  G  T  Y  M  L  A  K  Y  A  G  L  P  V  D  P  Q  S  S  G   -

              GTCCTGATGGTCCTCGTGTTCGGTGTCGGCACCGACTACGCCCTGCTGCTCATCGCCCGC
      44620   +---------+---------+---------+---------+---------+---------  44679
              CAGGACTACCAGGAGCACAAGCCACAGCCGTGGCTGATGCGGGACGACGAGTAGCGGGCG
      27         V  L  M  V  L  V  F  G  V  G  T  D  Y  A  L  L  L  I  A  R   -

              TACCGTGAGGAACTGCGCCGCGAGCAGGACCGGCACGTGGCCATGAAGACCGCGTTGCGA
      44680   +---------+---------+---------+---------+---------+---------  44739
              ATGGCACTCCTTGACGCGGCGCTCGTCCTGGCCGTGCACCGGTACTTCTGGCGCAACGCT
      27         Y  R  E  E  L  R  R  E  Q  D  R  H  V  A  M  K  T  A  L  R   -

              CGGTCGGGCCCGGCCATCCTGGCCTCGGCCGGCACCATCGCCATCGGCCTCGTCTGCCTG
```

56

```
                          ●                                    ●

                GCCAGCCCGGGCCGGTAGGACCGGAGCCGGCCGTGGTAGCGGTAGCCGGAGCAGACGGAC
   44740  +---------+---------+---------+---------+--------+---------  44799
                GTCGGGCCCGGCCATCCTGGCCTCGGCCGGCACCATCGCCATCGGCCTCGTCTGCCTG
   27          R  S  G  P  A  I  L  A  S  A  G  T  I  A  I  G  L  V  C  L    -

                GTCCTCGCGGACGTCAACTCCTCCCGCTCCATGGGCCTGGTCGGCGCGATCGGCGTGGTC
   44800  +---------+---------+---------+---------+--------+---------  44859
                CAGGAGCGCCTGCAGTTGAGGAGGGCGAGGTACCCGGACCAGCCGCGCTAGCCGCACCAG
   27           V  L  A  D  V  N  S  S  R  S  M  G  L  V  G  A  I  G  V  V    -

                TGCGCCCTCCTCGCCATGGTCACGATCCTGCCCGCGCTGCTGGTCATCCTGGGCCGCTGG
   44860  +---------+---------+---------+---------+--------+---------  44919
                ACGCGGGAGGAGCGGTACCAGTGCTAGGACGGGCGCGACGACCAGTAGGACCCGGCGACC
   27           C  A  L  L  A  M  V  T  I  L  P  A  L  L  V  I  L  G  R  W    -

                GTGTTCTGGCCCTTCGTTCCCCGCTGGACGCCGGAGTCGGCCGCGGCCCCCGAGGCACCG
   44920  +---------+---------+---------+---------+--------+---------  44979
                CACAAGACCGGGAAGCAAGGGGCGACCTGCGGCCTCAGCCGGCGCCGGGGGCTCCGTGGC
   27           V  F  W  P  F  V  P  R  W  T  P  E  S  A  A  A  P  E  A  P    -

                GCGTCCCACAGCCGCTGGGAGCGCATCGGCTCCGTCACGGCCGCCCGGCCGCGCCGCGCC
   44980  +---------+---------+---------+---------+--------+---------  45039
                CGCAGGGTGTCGGCGACCCTCGCGTAGCCGAGGCAGTGCCGGCGGGCCGGCGCGGCGCGG
   27           A  S  H  S  R  W  E  R  I  G  S  V  T  A  A  R  P  R  R  A    -

                TGGGTGCTGTCCTTGGCCGCGACGGGGCTTCTCGCCCTCAGTTCCCTCGGCCTCGACATG
   45040  +---------+---------+---------+---------+--------+---------  45099
                ACCCACGACAGGAACCGGCGCTGCCCCGAAGAGCGGGAGTCAAGGGAGCCGGAGCTGTAC
   27           W  V  L  S  L  A  A  T  G  L  L  A  L  S  S  L  G  L  D  M    -

                GGACTCACCCAGAGCGAACTGCTCCAGACGAAGCCCGAGTCCGTCGTCGCCCAGGAGCGG
   45100  +---------+---------+---------+---------+--------+---------  45159
                CCTGAGTGGGTCTCGCTTGACGAGGTCTGCTTCGGGCTCAGGCAGCAGCGGGTCCTCGCC
   27           G  L  T  Q  S  E  L  L  Q  T  K  P  E  S  V  V  A  Q  E  R    -

                ATCTCCGCCCACTACCCGTCCGGCTCCTCCGACCCCGCCACCGTCGTCGCACCCAGCGCG
   45160  +---------+---------+---------+---------+--------+---------  45219
                TAGAGGCGGGTGATGGGCAGGCCGAGGAGGCTGGGGCGGTGGCAGCAGCGTGGGTCGCGC
   27           I  S  A  H  Y  P  S  G  S  S  D  P  A  T  V  V  A  P  S  A    -

                GACGTGGCCGAGGTCCGCCGGGCCGCCGAGGGGACCGACGGAGTGGTCTCCGTCCAGGAC
   45220  +---------+---------+---------+---------+--------+---------  45279
                CTGCACCGGCTCCAGGCGGCCCGGCGGCTCCCCTGGCTGCCTCACCAGAGGCAGGTCCTG
   27           D  V  A  E  V  R  R  A  A  E  G  T  D  G  V  V  S  V  Q  D    -

                GGCCCCACCACTCCCGACGGAGAGCTGACCATGCTGTCCGTGGTGCTGAAGGACGTTCCC
   45280  +---------+---------+---------+---------+--------+---------  45339
                CCGGGGTGGTGAGGGCTGCCTCTCGACTGGTACGACAGGCACCACGACTTCCTGCAAGGG
   27           G  P  T  T  P  D  G  E  L  T  M  L  S  V  V  L  K  D  V  P    -

                GACAGCAGCGGGGCCAAGGACACCATCGATGCACTGCGGGACAACACGGATGCTCTCGTG
   45340  +---------+---------+---------+---------+--------+---------  45399
                CTGTCGTCGCCCCGGTTCCTGTGGTAGCTACGTGACGCCCTGTTGTGCCTACGAGAGCAC
   27           D  S  S  G  A  K  D  T  I  D  A  L  R  D  N  T  D  A  L  V    -

                GGGGGTACGACGGCCCAGAGCCTGGACACCCAGCGCGCCTCGGTCCGTGACCTCTGGGTC
   45400  +---------+---------+---------+---------+--------+---------  45459
                CCCCCATGCTGCCGGGTCTCGGACCTGTGGGTCGCGCGGAGCCAGGCACTGGAGACCCAG
   27           G  G  T  T  A  Q  S  L  D  T  Q  R  A  S  V  R  D  L  W  V    -

                ACCGTCCCCGCGGTCCTGCTGGTGGTCCTGCTCGTCCTGATCTGGCTGCTGCGCTCGGTC
   45460  +---------+---------+---------+---------+--------+---------  45519
                TGGCAGGGGCGCCAGGACGACCACCAGGACGAGCAGGACTAGACCGACGACGCGAGCCAG
   27           T  V  P  A  V  L  L  V  V  L  L  V  L  I  W  L  L  R  S  V    -

                ACCGGACCGCTGATCATGCTCGGCACCGTGGTCGTGTCGTTCTTCGCGGCCCTGGGGGCG
   45520  +---------+---------+---------+---------+--------+---------  45579
```

57

```
                 TGGCCTGGCGACTAGTACGAGCCGTGGCACCAGCACAGCAAGAAGCGCCGGGACCCCCGC
        27        T  G  P  L  I  M  L  G  T  V  V  V  S  F  F  A  A  L  G  A   -

                 TCCAACCTGCTCTTCGAGTACGTGATGGGGCACGCCGGCGTCGACTGGTCGGTGCCGCTT
        45580    +---------+---------+---------+---------+---------+---------   45639
                 AGGTTGGACGAGAAGCTCATGCACTACCCCGTGCGGCCGCAGCTGACCAGCCACGGCGAA
        27        S  N  L  L  F  E  Y  V  M  G  H  A  G  V  D  W  S  V  P  L   -

                 CTCGGGTTCGTGTACCTGGTCGCCCTCGGAATCGACTACAACATCTTCCTCATGCACCGG
        45640    +---------+---------+---------+---------+---------+---------   45699
                 GAGCCCAAGCACATGGACCAGCGGGAGCCTTAGCTGATGTTGTAGAAGGAGTACGTGGCC
        27        L  G  F  V  Y  L  V  A  L  G  I  D  Y  N  I  F  L  M  H  R   -

                 GTGAAGGAGGAGGTCGCTCTGCACGGCCATGCCAAGGGCGTGCTCACCGGCCTGACCACC
        45700    +---------+---------+---------+---------+---------+---------   45759
                 CACTTCCTCCTCCAGCGAGACGTGCCGGTACGGTTCCCGCACGAGTGGCCGGACTGGTGG
        27        V  K  E  E  V  A  L  H  G  H  A  K  G  V  L  T  G  L  T  T   -

                 ACCGGGGGCGTCATCACCAGTGCCGGCGTGGTCCTGGCCGCGACGTTCGCCGTCATCGCC
        45760    +---------+---------+---------+---------+---------+---------   45819
                 TGGCCCCCGCAGTAGTGGTCACGGCCGCACCAGGACCGGCGCTGCAAGCGGCAGTAGCGG
        27        T  G  G  V  I  T  S  A  G  V  V  L  A  A  T  F  A  V  I  A   -

                 ACACTGCCGCTGGTCCCGATGGCCCAGATGGGTGTCGTGGTCGGCCTGGGCATTCTGCTG
        45820    +---------+---------+---------+---------+---------+---------   45879
                 TGTGACGGCGACCAGGGCTACCGGGTCTACCCACAGCACCAGCCGGACCCGTAAGACGAC
        27        T  L  P  L  V  P  M  A  Q  M  G  V  V  V  G  L  G  I  L  L   -

                 GACACCTTCCTCGTCCGGACGATTCTTCTGCCGGCCCTGGCGCTCGATCTGGGGCCCCGG
        45880    +---------+---------+---------+---------+---------+---------   45939
                 CTGTGGAAGGAGCAGGCCTGCTAAGAAGACGGCCGGGACCGCGAGCTAGACCCCGGGGCC
        27        D  T  F  L  V  R  T  I  L  L  P  A  L  A  L  D  L  G  P  R   -

                 TTCTGGTGGCCGGGCGCGCTGTCGAAGACGTCCGGGGGACCGGCCCCCGTCCGCGAGGAC
        45940    +---------+---------+---------+---------+---------+---------   45999
                 AAGACCACCGGCCCGCGCGACAGCTTCTGCAGGCCCCCTGGCCGGGGGCAGGCGCTCCTG
        27        F  W  W  P  G  A  L  S  K  T  S  G  G  P  A  P  V  R  E  D   -

                 CGCACGTCCCAGCCCGTGGGCTGAGACCCGTCCCGACGAGACCCGTACGGCGGGCGGCCG
        46000    +---------+---------+---------+---------+---------+---------   46059
                 GCGTGCAGGGTCGGGCACCCGACTCTGGGCAGGGCTGCTCTGGGCATGCCGCCCGCCGGC
        27        R  T  S  Q  P  V  G  * (ORF27)

                 GTTCCCCCGGGCCGTACGACTGAGCAACCCAGAAGATGGGCCGCCCGCGACCAGGCGTCA
        46060    +---------+---------+---------+---------+---------+---------   46119
                 CAAGGGGGCCCGGCATGCTGACTCGTTGGGTCTTCTACCCGGCGGGCGCTGGTCCGCAGT

                 CGATGGTGGCCCACCGGCCGCAGGCCGATCTCCCGGAAGGAAGCGCCGTGTTGGGCGATG
        46120    +---------+---------+---------+---------+---------+---------   46179
                 GCTACCACCGGGTGGCCGGCGTCCGGCTAGAGGGCCTTCCTTCGCGGCACAACCCGCTAC
        28                                            (ORF28)  V  L  G  D  E  -

                 AGGACGGCAAGGCCGCCGAGCTGTGGTCGATGGCGAACCTGGGTACACCGATGGCCGTGC
        46180    +---------+---------+---------+---------+---------+---------   46239
                 TCCTGCCGTTCCGGCGGCTCGACACCAGCTACCGCTTGGACCCATGTGGCTACCGGCACG
        28        D  G  K  A  A  E  L  W  S  M  A  N  L  G  T  P  M  A  V  R   -

                 GCGTCGCGGCGACCCTGCGCATCGCCGACCACATCACGGCCGGAGCGCACACCGCCGGCG
        46240    +---------+---------+---------+---------+---------+---------   46299
                 CGCAGCGCCGCTGGGACGCGTAGCGGCTGGTGTAGTGCCGGCCTCGCGTGTGGCGGCCGC
        28        V  A  A  T  L  R  I  A  D  H  I  T  A  G  A  H  T  A  G  E   -

                 AAATCGCCGAAGCGGCCGCCGTGCACGAGGAATCCCTCGACCGGCTGCTGCGCTACCTCA
        46300    +---------+---------+---------+---------+---------+---------   46359
                 TTTAGCGGCTTCGCCGGCGGCACGTGCTCCTTAGGGAGCTGGCCGACGACGCGATGGAGT
        28        I  A  E  A  A  A  V  H  E  E  S  L  D  R  L  L  R  Y  L  T   -
```

58

```
                CCGTCCGGGGCCTGCTGGACCGTGACGGGCTCGGCCGGTACACGCTGACCCCCCTGGGCC
       46360    +---------+---------+---------+---------+---------+---------    46419
                GGCAGGCCCCGGACGACCTGGCACTGCCCGAGCCGGCCATGTGCGACTGGGGGGACCCGG
       28          V  R  G  L  L  D  R  D  G  L  G  R  Y  T  L  T  P  L  G  R -

                GGCCGCTGTGCGAGGACCACCCCGCCGGCGTCCGGGCCTGGTTCGACATGGAGGGAGCGG
       46420    +---------+---------+---------+---------+---------+---------    46479
                CCGGCGACACGCTCCTGGTGGGGCGGCCGCAGGCCCGGACCAAGCTGTACCTCCCTCGCC
       28          P  L  C  E  D  H  P  A  G  V  R  A  W  F  D  M  E  G  A  G -

                GGCGGGGCGAGCTGTCGTTCGTCGACCTGCTGCACAGCGTACGGACCGGGAAGGCCGCCT
       46480    +---------+---------+---------+---------+---------+---------    46539
                CCGCCCCGCTCGACAGCAAGCAGCTGGACGACGTGTCGCATGCCTGGCCCTTCCGGCGGA
       28          R  G  E  L  S  F  V  D  L  L  H  S  V  R  T  G  K  A  A  F -

                TCCCCCTGCGCTACGGCCGCCCCTTCTGGGAGGACCTGGCGGAGGACCCCCGCCGCGCGG
       46540    +---------+---------+---------+---------+---------+---------    46599
                AGGGGGACGCGATGCCGGCGGGGAAGACCCTCCTGGACCGCCTCCTGGGGGCGGCGCGCC
       28          P  L  R  Y  G  R  P  F  W  E  D  L  A  E  D  P  R  R  A  E -

                AGTCCTTCAACCGGCTGCTCGGCCAGGACGTCGCCACTCGCGCCCCGGCCGTGGTGGCCG
       46600    +---------+---------+---------+---------+---------+---------    46659
                TCAGGAAGTTGGCCGACGAGCCGGTCCTGCAGCGGTGAGCGCGGGGCCGGCACCACCGGC
       28          S  F  N  R  L  L  G  Q  D  V  A  T  R  A  P  A  V  V  A  G -

                GCTTCGACTGGGCGAGCACCGGTCATGTCATCGACCTCGGAGGCGGCGACGGCTCCCTGC
       46660    +---------+---------+---------+---------+---------+---------    46719
                CGAAGCTGACCCGCTCGTGGCCAGTACAGTAGCTGGAGCCTCCGCCGCTGCCGAGGGACG
       28          F  D  W  A  S  T  G  H  V  I  D  L  G  G  G  D  G  S  L  L -

                TGACCGCACTGCTGACCGCCTGTCCGTCACTGCGCGGCACGGTCCTGGACCTGCCCGAAG
       46720    +---------+---------+---------+---------+---------+---------    46779
                ACTGGCGTGACGACTGGCGGACAGGCAGTGACGCGCCGTGCCAGGACCTGGACGGGCTTC
       28          T  A  L  L  T  A  C  P  S  L  R  G  T  V  L  D  L  P  E  A -

                CGGTGCAGCGTGCCAAGGAGTCGTTCGCCGTGTCCGGACTGGACGACCGGGCGAACGCGG
       46780    +---------+---------+---------+---------+---------+---------    46839
                GCCACGTCGCACGGTTCCTCAGCAAGCGGCACAGGCCTGACCTGCTGGCCCGCTTGCGCC
       28          V  Q  R  A  K  E  S  F  A  V  S  G  L  D  D  R  A  N  A  V -

                TCGCGGGCAGCTTCTTCGACGCCCTCCCCGCCGGCGCGGGCGCCTACGTCCTGTCCCTGG
       46840    +---------+---------+---------+---------+---------+---------    46899
                AGCGCCCGTCGAAGAAGCTGCGGGAGGGGCGGCCGCGCCCGCGGATGCAGGACAGGGACC
       28          A  G  S  F  F  D  A  L  P  A  G  A  G  A  Y  V  L  S  L  V -

                TCCTGCACGACTGGGACGACGAGGCGTCCGTCGCGATCCTGCGGCGCTGCGCCGAGGCGG
       46900    +---------+---------+---------+---------+---------+---------    46959
                AGGACGTGCTGACCCTGCTGCTCCGCAGGCAGCGCTAGGACGCCGCGACGCGGCTCCGCC
       28          L  H  D  W  D  D  E  A  S  V  A  I  L  R  R  C  A  E  A  A -

                CGGGGCAGACGGGATCGGTGTTCGTCATCGAGTCGACCGGCTCGGCGGGGGACGCCCCGC
       46960    +---------+---------+---------+---------+---------+---------    47019
                GCCCCGTCTGCCCTAGCCACAAGCAGTAGCTCAGCTGGCCGAGCCGCCCCCTGCGGGGCG
       28          G  Q  T  G  S  V  F  V  I  E  S  T  G  S  A  G  D  A  P  H -

                ACACAGGTATGGACCTGCGCATGCTGTGCATCTACGGAGCCAAGGAGCGCCGCGTGGAGG
       47020    +---------+---------+---------+---------+---------+---------    47079
                TGTGTCCATACCTGGACGCGTACGACACGTAGATGCCTCGGTTCCTCGCGGCGCACCTCC
       28          T  G  M  D  L  R  M  L  C  I  Y  G  A  K  E  R  R  V  E  E -

                AGTTCGAGGAACTCGCCGGCCGGGCCGGGCTCCGGGTCGTCGCCGTCCACCCCGCGGGCC
       47080    +---------+---------+---------+---------+---------+---------    47139
                TCAAGCTCCTTGAGCGGCCGGCCCGGCCCGAGGCCCAGCAGCGGCAGGTGGGGCGCCCGG
       28          F  E  E  L  A  G  R  A  G  L  R  V  V  A  V  H  P  A  G  P -
```

59

```
        CTTCCGCGATCATCCAGATGTCCGCGGTCTGACCGCCCGGAGCCCCGGCCCATCGCGGCG
 47140  +---------+---------+---------+---------+---------+---------  47199
        GAAGGCGCTAGTAGGTCTACAGGCGCCAGACTGGCGGGCCTCGGGGCCGGGTAGCGCCGC
  28       S  A  I  I  Q  M  S  A  V  *  (ORF28)


        CGGGCCACGGCAGACAAGGAGAGAGCGTATGGCCGGCCTGGTCATGTCGCCGGTGGAGGC
 47200  +---------+---------+---------+---------+---------+---------  47259
        GCCCGGTGCCGTCTGTTCCTCTCTCGCATACCGGCCGGACCAGTACAGCGGCCACCTCCG
                        (ORF29)  M  A  G  L  V  M  S  P  V  E  A   -


        GCTCGACGCGCTGGGCACGGTGCAGGGGCGTCAGGACCCCTATCCCTTCTACGAGGCGAT
 47260  +---------+---------+---------+---------+---------+---------  47319
        CGAGCTGCGCGACCCGTGCCACGTCCCCGCAGTCCTGGGGATAGGGAAGATGCTCCGCTA
  29       L  D  A  L  G  T  V  Q  G  R  Q  D  P  Y  P  F  Y  E  A  I  -


        CCGCGCGCACGGGCAGGCGGTCCCCACGAAGCCCGGCCGCTTCGTGGTGGTCGGCCACGA
 47320  +---------+---------+---------+---------+---------+---------  47379
        GGCGCGCGTGCCCGTCCGCCAGGGGTGCTTCGGGCCGGCGAAGCACCACCAGCCGGTGCT
  29       R  A  H  G  Q  A  V  P  T  K  P  G  R  F  V  V  V  G  H  D  -


        CGCGTGCGACCGGGCGCTGCGGGAACCGGCCCTGCGCGTCCAGGACGCCAGGAGCTACGA
 47380  +---------+---------+---------+---------+---------+---------  47439
        GCGCACGCTGGCCCGCGACGCCCTTGGCCGGGACGCGCAGGTCCTGCGGTCCTCGATGCT
  29       A  C  D  R  A  L  R  E  P  A  L  R  V  Q  D  A  R  S  Y  D  -


        CGTCGTCTTCCCCTCGTGGCGGTCGCACTCCTCGGTCCGGGGGTTCACCAGCTCCATGCT
 47440  +---------+---------+---------+---------+---------+---------  47499
        GCAGCAGAAGGGGAGCACCGCCAGCGTGAGGAGCCAGGCCCCCAAGTGGTCGAGGTACGA
  29       V  V  F  P  S  W  R  S  H  S  S  V  R  G  F  T  S  S  M  L  -


        CTACAGCAACCCGCCCGATCACGGCCGGTTGCGCCAGGTGGTGAGCTTCGCGTTCACCCC
 47500  +---------+---------+---------+---------+---------+---------  47559
        GATGTCGTTGGGCGGGCTAGTGCCGGCCAACGCGGTCCACCACTCGAAGCGCAAGTGGGG
  29       Y  S  N  P  P  D  H  G  R  L  R  Q  V  V  S  F  A  F  T  P  -


        GCCCAAGGTGCGCCGGATGCACGGGGTGATCGAGGACATGACCGACCGGCTCCTCGACCG
 47560  +---------+---------+---------+---------+---------+---------  47619
        CGGGTTCCACGCGGCCTACGTGCCCCACTAGCTCCTGTACTGGCTGGCCGAGGAGCTGGC
  29       P  K  V  R  R  M  H  G  V  I  E  D  M  T  D  R  L  L  D  R  -


        GATGGCCCGGCTCGGCTCCGGCGGCTCCCCGGTCGACCTCATAGCCGAGTTCGCCGCCCG
 47620  +---------+---------+---------+---------+---------+---------  47679
        CTACCGGGCCGAGCCGAGGCCGCCGAGGGGCCAGCTGGAGTATCGGCTCAAGCGGCGGGC
  29       M  A  R  L  G  S  G  G  S  P  V  D  L  I  A  E  F  A  A  R  -


        GCTGCCCGTCGCGGTGATCAGCGAGATGATCGGCTTTCCGGCGAAGGACCAGGTGTGGTT
 47680  +-----,---+---------+---------+---------+---------+---------  47739
        CGACGGGCAGCGCCACTAGTCGCTCTACTAGCCGAAAGGCCGCTTCCTGGTCCACACCAA
  29       L  P  V  A  V  I  S  E  M  I  G  F  P  A  K  D  Q  V  W  F  -


        CCGCGACATGGCCTCCCGGGTCGCCGTGGCGACGGACGGTTTCACCGACCCCGGCGCGCT
 47740  +---------+---------+---------+---------+---------+---------  47799
        GGCGCTGTACCGGAGGGCCCAGCGGCACCGCTGCCTGCCAAAGTGGCTGGGGCCGCGCGA
  29       R  D  M  A  S  R  V  A  V  A  T  D  G  F  T  D  P  G  A  L  -


        CACGGGGGCCGACGCCGCCATGGACGAGATGAGCGCCTACTTCGACGACCTCCTGGACCG
 47800  +---------+---------+---------+---------+---------+---------  47859
        GTGCCCCCGGCTGCGGCGGTACCTGCTCTACTCGCGGATGAAGCTGCTGGAGGACCTGGC
  29       T  G  A  D  A  A  M  D  E  M  S  A  Y  F  D  D  L  L  D  R  -


        TCGCCGCCGCACCCCGGCCGACGACCTGGTCACCCTGCTCGCCGAGGCCCACGACGGCTC
 47860  +---------+---------+---------+---------+---------+---------  47919
        AGCGGCGGCGTGGGGCCGGCTGCTGGACCAGTGGGACGAGCGGCTCCGGGTGCTGCCGAG
  29       R  R  R  T  P  A  D  D  L  V  T  L  L  A  E  A  H  D  G  S  -


        CCCCGGGCGCCTGGACCACGACGAACTGATGGGCACCATGATGGTGCTGCTCACAGCCGG
```

60

```
                    ●                              ●

              GGGGCCCGCGGACCTGGTGCTGCTTGACTACCCGTGGTACTACCACGACGAGTGTCGGCC
       47920  +---------+---------+---------+---------+---------+--------- 47979
              CCCCGGGCGCCTGGACCACGACGAACTGATGGGCACCATGATGGTGCTGCTCACAGCCGG
        29        P  G  R  L  D  H  D  E  L  M  G  T  M  M  V  L  L  T  A  G  -

              GTTCGAGACCACGAGCTTTCTGATCGGCCACGGGGCGATGATCGCCCTCGAACAACGGGC
       47980  +---------+---------+---------+---------+---------+--------- 48039
              CAAGCTCTGGTGCTCGAAAGACTAGCCGGTGCCCCGCTACTAGCGGGAGCTTGTTGCCCG
        29        F  E  T  T  S  F  L  I  G  H  G  A  M  I  A  L  E  Q  R  A  -

              GCACGCGGCCCGGCTGCGGGCCGAACCCGACTTCGCCGACGGCTACGTCGAGGAGATCCT
       48040  +---------+---------+---------+---------+---------+--------- 48099
              CGTGCGCCGGGCCGACGCCCGGCTTGGGCTGAAGCGGCTGCCGATGCAGCTCCTCTAGGA
        29        H  A  A  R  L  R  A  E  P  D  F  A  D  G  Y  V  E  E  I  L  -

              CAGGTTCGAGCCGCCGGTCCACGTCACCAGCCGGTGGGCTGCCGAGGACCTCGACCTGCT
       48100  +---------+---------+---------+---------+---------+--------- 48159
              GTCCAAGCTCGGCGGCCAGGTGCAGTGGTCGGCCACCCGACGGCTCCTGGAGCTGGACGA
        29        R  F  E  P  P  V  H  V  T  S  R  W  A  A  E  D  L  D  L  L  -

              GGGCCTGTCCGTACCGGCGGGCTCCAAGCTGGTCCTGATCCTGGCCGCCGCGAATCGCGA
       48160  +---------+---------+---------+---------+---------+--------- 48219
              CCCGGACAGGCATGGCCGCCCGAGGTTCGACCAGGACTAGGACCGGCGGCGCTTAGCGCT
        29        G  L  S  V  P  A  G  S  K  L  V  L  I  L  A  A  A  N  R  D  -

              TCCCGGCCGCTACCCCGAGCCCGGCCGCTTCGACCCCGACCGCTACGCGCCCCGGCCGGG
       48220  +---------+---------+---------+---------+---------+--------- 48279
              AGGGCCGGCGATGGGGCTCGGGCCGGCGAAGCTGGGGCTGGCGATGCGCGGGGCCGGCCC
        29        P  G  R  Y  P  E  P  G  R  F  D  P  D  R  Y  A  P  R  P  G  -

              CGGGCCGGAGGCCACCAGACCGCTGAGCTTCGGCGCGGGCGGCCACTTCTGCCTCGGCGC
       48280  +---------+---------+---------+---------+---------+--------- 48339
              GCCCGGCCTCCGGTGGTCTGGCGACTCGAAGCCGCGCCCGCCGGTGAAGACGGAGCCGCG
        29        G  P  E  A  T  R  P  L  S  F  G  A  G  G  H  F  C  L  G  A  -

              TCCGCTGGCGCGGCTGGAAGCCCGGATCGCGCTGCCGCGTCTGCTGCGCCGCTTCCCGGA
       48340  +---------+---------+---------+---------+---------+--------- 48399
              AGGCGACCGCGCCGACCTTCGGGCCTAGCGCGACGGCGCAGACGACGCGGCGAAGGGCCT
        29        P  L  A  R  L  E  A  R  I  A  L  P  R  L  L  R  R  F  P  D  -

              CCTGGCCGTGTCCGAGCCCCCCGTCTACCGCGACCGCTGGGTCGTCCGCGGCCTCGAAAC
       48400  +---------+---------+---------+---------+---------+--------- 48459
              GGACCGGCACAGGCTCGGGGGGCAGATGGCGCTGGCGACCCAGCAGGCGCCGGAGCTTTG
        29        L  A  V  S  E  P  P  V  Y  R  D  R  W  V  V  R  G  L  E  T  -

              CTTTCCCGTGACCCTCGGGTCCTGAGCCCCCGCCGGCCGGAACACGTGACCGTCCCGGCC
       48460  +---------+---------+---------+---------+---------+--------- 48519
              GAAAGGGCACTGGGAGCCCAGGACTCGGGGGCGGCCGGCCTTGTGCACTGGCAGGGCCGG
        29        F  P  V  T  L  G  S  *  (ORF29)

              GGCGGGTGCGCGCCCTCTCAGACGTACAGGGTGTTGGGCCCCTGACCACACAGCACCCGG
       48520  +---------+---------+---------+---------+---------+--------- 48579
              CCGCCCACGCGCGGGAGAGTCTGCATGTCCCACAACCCGGGGACTGGTGTGTCGTGGGCC

              CCGTACAGCTCCAGGTTGGTGCTCGGGTTCATGCAGGTGCAGCGTGATGCTCTGGGCATC
       48580  +---------+---------+---------+---------+---------+--------- 48639
              GGCATGTCGAGGTCCAACCACGAGCCCAAGTACGTCCACGTCGCACTACGAGACCCGTAG
        30                                 (ORF30)*  A  P  A  A  H  H  E  P  C

              GCTGCACGCGCTGGATCGGGACGTCGTTGTAGATCGAGGACCCGCCGCTCGCCTGGGCGA
       48640  +---------+---------+---------+---------+---------+--------- 48699
              CGACGTGCGCGACCTAGCCCTGCAGCAACATCTAGCTCCTGGGCGGCGAGCGGACCCGCT
        30        R  Q  V  R  Q  I  P  V  D  N  Y  I  S  S  G  G  S  A  Q  A  -

              GGATGTCCACCGACTCCTTGCCCAGTCGGCACGCCCGCCCCAGCAGGCCGCGGCACAGCA
       48700  +---------+---------+---------+---------+---------+--------- 48759
              CCTACAGGTGGCTGAGGAACGGGTCAGCCGTGCGGGCGGGGTCGTCCGGCGCCGTGTCGT
```

61

```
                    L   I   D   V   S   E   K   G   L   R   C   A   R   G   L   L   G   R   C   L   -

              CCCGCTCCTCCAGCGTCCAGGCCTCGCCCGAAGCCCCCTTGGAGTCGACGAGGTCGGCCA
       48760  +---------+---------+---------+---------+---------+---------  48819
              GGGCGAGGAGGTCGCAGGTCCGGAGCGGGCTTCGGGGGAACCTCAGCTGCTCCAGCCGGT
       30       V   R   E   E   L   T   W   A   E   G   S   A   G   K   S   D   V   L   D   A   -

              GCCGATGGGCGTGGAACCGTGCCTCGTCGGCCAGCAGGGTCGCCTCGCCGAGCTGCAGGT
       48820  +---------+---------+---------+---------+---------+---------  48879
              CGGCTACCCGCACCTTGGCACGGAGCAGCCGGTCGTCCCAGCGGAGCGGCTCGACGTCCA
       30       L   R   H   A   H   F   R   A   E   D   A   L   L   T   A   E   G   L   Q   L   -

              GGGTGATCGGCGCCGAGCCCTGCTCCTCGTACTCGGTGTAGGTGATCTTGCGGCCGGGCA
       48880  +---------+---------+---------+---------+---------+---------  48939
              CCCACTAGCCGCGGCTCGGGACGAGGAGCATGAGCCACATCCACTAGAACGCCGGCCCGT
       30       H   T   I   P   A   S   G   Q   E   E   Y   E   T   Y   T   I   K   R   G   P   -

              GCCTCCCGCGGAAGACGTCCTGAGCGGCCGCGGCCAGTCCGGTCATGGTGCCGACCGACG
       48940  +---------+---------+---------+---------+---------+---------  48999
              CGGAGGGCGCCTTCTGCAGGACTCGCCGGCGCCGGTCAGGCCAGTACCACGGCTGGCTGC
       30       L   R   G   R   F   V   D   Q   A   A   A   A   L   G   T   M   T   G   V   S   -

              AGGCCGAGGCCACGGCCAGCATCGGCGCCCGGAACATCGGTGATCCGGCGTTGAGTTCGG
       49000  +---------+---------+---------+---------+---------+---------  49059
              TCCGGCTCCGGTGCCGGTCGTAGCCGCGGGCCTTGTAGCCACTAGGCCGCAACTCAAGCC
       30       S   A   S   A   V   A   L   M   P   A   R   F   M   P   S   G   A   N   L   E   -

              AGGCGTACTGCTGCTGGAGCACCGCGCCCAGCGGAAGGACGCGCTCCTGGGGAACGAAGA
       49060  +---------+---------+---------+---------+---------+---------  49119
              TCCGCATGACGACGACCTCGTGGCGCGGGTCGCCTTCCTGCGCGAGGACCCCTTGCTTCT
       30       S   A   Y   Q   Q   Q   L   V   A   G   L   P   L   V   R   E   Q   P   V   F   -

              CGTCCGCGGCGATGGTGCTGACGCTTCCCGAGCCCCGGAGCCCCGAGGTGTGCCAGTCGT
       49120  +---------+---------+---------+---------+---------+---------  49179
              GCAGGCGCCGCTACCACGACTGCGAAGGGCTCGGGGCCTCGGGGCTCCACACGGTCAGCA
       30       V   D   A   A   I   T   S   V   S   G   S   G   R   L   G   S   T   H   W   D   -

              CGACGATCTGCAGCTGGTCGGTCGGCACCAGGGCCATCACGGGCTGCATGCCGCCGTCGG
       49180  +---------+---------+---------+---------+---------+---------  49239
              GCTGCTAGACGTCGACCAGCCAGCCGTGGTCCCGGTAGTGCCCGACGTACGGCGGCAGCC
       30       D   V   I   Q   L   Q   D   T   P   V   L   A   M   V   P   Q   M   G   G   D   -

              GGGTCGGTGAGACGGCGATCAGAACCTGCCAGTGACTGTGCCAGGCACCGCTGATGAAGC
       49240  +---------+---------+---------+---------+---------+---------  49299
              CCCAGCCACTCTGCCGCTAGTCTTGGACGGTCACTGACACGGTCCGTGGCGACTACTTCG
       30       P   T   P   S   V   A   I   L   V   Q   W   H   S   H   W   A   G   S   I   F   -

              CCCACTTGCCGTTCACTACGACACCGCCGTCGACCGGGGCCGCCATGCCGCCGGGACTGA
       49300  +---------+---------+---------+---------+---------+---------  49359
              GGGTGAACGGCAAGTGATGCTGTGGCGGCAGCTGGCCCCGGCGGTACGGCGGCCCTGACT
       30       G   W   K   G   N   V   V   V   G   G   D   V   P   A   A   M   G   G   P   S   -

              GGGTGCCGGAGACCCGGACATCCGGCCGGGAGAACACCTCGTCCTGCACGTGGTCGGGGA
       49360  +---------+---------+---------+---------+---------+---------  49419
              CCCACGGCCTCTGGGCCTGTAGGCCGGCCCTCTTGTGGAGCAGGACGTGCACCAGCCCCT
       30       L   T   G   S   V   R   V   D   P   R   S   F   V   E   D   Q   V   H   D   P   -

              AGAGGCCCGCCATCCAGGTGGGTATCCACCACACCGAGGCCGTCCAGGCGGCCGATCCGT
       49420  +---------+---------+---------+---------+---------+---------  49479
              TCTCCGGGCGGTAGGTCCACCCATAGGTGGTGTGGCTCCGGCAGGTCCGCCGGCTAGGCA
       30       F   L   G   A   M   W   T   P   I   W   W   V   S   A   T   W   A   A   S   G   -

              CGCCGCGCGCCAGCTCGGCGGCCACGTCCACCAGGGTGCGGGCGTCGGACTCGAAGCCGC
       49480  +---------+---------+---------+---------+---------+---------  49539
              GCGGCGCGCGGTCGAGCCGCCGGTGCAGGTGGTCCCACGCCCGCAGCCTGAGCTTCGGCG
       30       D   G   R   A   L   E   A   A   V   D   V   L   T   R   A   D   S   E   F   G   -
```

62

```
          CGTAACGGGCCGGCACGCGCATGCGGAAGATCCCGGCTTCGGCCATCGCCTCGACCGACT
49540     +---------+---------+---------+---------+---------+--------- 49599
          GCATTGCCCGGCCGTGCGCGTACGCCTTCTAGGGCCGAAGCCGGTAGCGGAGCTGGCTGA
30          G  Y  R  A  P  V  R  M  R  F  I  G  A  E  A  M  A  E  V  S   -

          CCTCGTGCAGCCGCCGGTTCTCCTCGGTCCAGGCCGCGTGGGACTGGAGCAGCGGCCTCA
49600     +---------+---------+---------+---------+---------+--------- 49659
          GGAGCACGTCGGCGGCCAAGAGGAGCCAGGTCCGGCGCACCCTGACCTCGTCGCCGGAGT
30          E  E  H  L  R  R  N  E  E  T  W  A  A  H  S  Q  L  L  P  R   -

          GCTTCGAGGCCCGTTCCACCAGTTCGGTACGGGCGGGCGTAGACGTCTGGTCCACTCGAT
49660     +---------+---------+---------+---------+---------+--------- 49719
          CGAAGCTCCGGGCAAGGTGGTCAAGCCATGCCCGCCCGCATCTGCAGACCAGGTGAGCTA
30          L  K  S  A  R  E  V  L  E  T  R  A  P  T  S  T  Q  D  V
(ORF30)
          CCTCCAGGAATCATGAGACGCCCTGTCCGCGGTATGCGGAAGCAGGCGTCTGCGCGCATC
49720     +---------+---------+---------+---------+---------+--------- 49779
          GGAGGTCCTTAGTACTCTGCGGGACAGGCGCCATACGCCTTCGTCCGCAGACGCGCGTAG

          GGTCAGGACGGCGTCGCCCTGCTCCCGCATGGTTCACCGAGTTCCGCGGACGTCGCATCT
49780     +---------+---------+---------+---------+---------+--------- 49839
          CCAGTCCTGCCGCAGCGGGACGAGGGCGTACCAAGTGGCTCAAGGCGCCTGCAGCGTAGA

          CCTTGATTGCCGGTCACCTACCCCGATGCCGATCGGGCTGGTGCGACAGCGCATCCCACG
49840     +---------+---------+---------+---------+---------+--------- 49899
          GGAACTAACGGCCAGTGGATGGGGCTACGGCTAGCCCGACCACGCTGTCGCGTAGGGTGC

          AGAAGTCCACGAACGGTCCGGGAAGCCAGAATGTGCTTCTCGGCCGGAGTCACGGCCGGC
49900     +---------+---------+---------+---------+---------+--------- 49959
          TCTTCAGGTGCTTGCCAGGCCCTTCGGTCTTACACGAAGAGCCGGCCTCAGTGCCGGCCG

          GCCGGCGCCCGTCGCCGGTCACGCCGGACCACGCCCGGACCGGTCATGGAGGCAGCCCAT
49960     +---------+---------+---------+---------+---------+--------- 50019
          CGGCCGCGGGCAGCGGCCAGTGCGGCCTGGTGCGGGCCTGGCCAGTACCTCCGTCGGGTA

          GAGTGACAACGACAGTCCGTCCCGGGTGCCGGCCGCGGTGGCACCCGCCACCGCGAAACC
50020     +---------+---------+---------+---------+---------+--------- 50079
          CTCACTGTTGCTGTCAGGCAGGGCCCACGGCCGGCGCCACCGTGGGCGGTGGCGCTTTGG

          GTCGGCCGGCACGGTCCTCGGCGCCGCGGTGGCTTCGCCCGCCGCCTACACCGCGGCGAC
50080     +---------+---------+---------+---------+---------+--------- 50139
          CAGCCGGCCGTGCCAGGAGCCGCGGCGCCACCGAAGCGGGCGGCGGATGTGGCGCCGCTG

          CGCCCAGGAAGCGGCGACCGCGCTGGTCCGCATGCTGATGGAACAGATGGTGCTCGGTCC
50140     +---------+---------+---------+---------+---------+--------- 50199
          GCGGGTCCTTCGCCGCTGGCGCGACCAGGCGTACGACTACCTTGTCTACCACGAGCCAGG

          CGGCGCGGTCGGTCCCGAGACCCGCGCGGACGGCCCGGCGCGGCGGACCGGCTCCGGCCA
50200     +---------+---------+---------+---------+---------+--------- 50259
          GCCGCGCCAGCCAGGGCTCTGGGCGCGCCTGCCGGGCCGCGCCGCCTGGCCGAGGCCGGT

          CGGCCCGGCGCCGCAGACCGGACCGGACGCGCCGGGCGAACCCCCGCCCACGTGGGCGCC
50260     +---------+---------+---------+---------+---------+--------- 50319
          GCCGGGCCGCGGCGTCTGGCCTGGCCTGCGCGGCCCGCTTGGGGGCGGGTGCACCCGCGG

          GAACCTCGACGACGGGAAGGTAGGAGGACGATGAGGCCGCTCGTTCGGGCAGTGCTGCGG
50320     +---------+---------+---------+---------+---------+--------- 50379
          CTTGGAGCTGCTGCCCTTCCATCCTCCTGCTACTCCGGCGAGCAAGCCCGTCACGACGCC
31                                      (ORF31) M  R  P  L  V  R  A  V  L  R   -

          GGTTCCCTGCGGCAGGTGAGGTACGTGGACGTGGTCTCCCCGCGCCGGGCGCGCTCCCTG
50380     +---------+---------+---------+---------+---------+--------- 50439
          CCAAGGGACGCCGTCCACTCCATGCACCTGCACCAGAGGGGCGCGGCCCGCGCGAGGGAC
31          G  S  L  R  Q  V  R  Y  V  D  V  V  S  P  R  R  A  R  S  L   -
```

63

```
              ●                              ●


        GTGGCGCGGGTGTACCGGGAGACCGAGGAGCAGTTCGGCGTGCTCGCGCCCCCCCTGGCC
  50440 +---------+---------+---------+---------+---------+---------  50499
        CACCGCGCCCACATGGCCCTCTGGCTCCTCGTCAAGCCGCACGAGCGCGGGGGGGACCGG
   31     V  A  R  V  Y  R  E  T  E  E  Q  F  G  V  L  A  P  P  L  A    -

        CTCCACTCGCCCGCCGCGGCGTCGCTGGCCGCGACGTGGCTCATGCTGCGGGAGACACTG
  50500 +---------+---------+---------+---------+---------+---------  50559
        GAGGTGAGCGGGCGGCGCCGCAGCGACCGGCGCTGCACCGAGTACGACGCCCTCTGTGAC
   31     L  H  S  P  A  A  A  S  L  A  A  T  W  L  M  L  R  E  T  L    -

        CTGGTCGACGGGCGGGTGAGCCGGGCGGTGAAGGAGACGGTCGCCACCGAGGTCTCCCGT
  50560 +---------+---------+---------+---------+---------+---------  50619
        GACCAGCTGCCCGCCCACTCGGCCCGCCACTTCCTCTGCCAGCGGTGGCTCCAGAGGGCA
   31     L  V  D  G  R  V  S  R  A  V  K  E  T  V  A  T  E  V  S  R    -

        GCCAACGACTGTCCGTACTGCGTCCAGGTCCATCAGGCGGTACTCGGGACACTGCCTCCG
  50620 +---------+---------+---------+---------+---------+---------  50679
        CGGTTGCTGACAGGCATGACGCAGGTCCAGGTAGTCCGCCATGAGCCCTGTGACGGAGGC
   31     A  N  D  C  P  Y  C  V  Q  V  H  Q  A  V  L  G  T  L  P  P    -

        GACGGCGGCCAGGCCGGGCTCCTGCGGTGGGTCCGGGAGGCAGGCCGACGGCCCGGCGGC
  50680 +---------+---------+---------+---------+---------+---------  50739
        CTGCCGCCGGTCCGGCCCGAGGACGCCACCCAGGCCCTCCGTCCGGCTGCCGGGCCGCCG
   31     D  G  G  Q  A  G  L  L  R  W  V  R  E  A  G  R  R  P  G  G    -

        GGTGCGGTGGGCGGCGGGCGGCCGCTTCCGTTCAGCGGTGAACAGGCACCGGAACTGTGC
  50740 +---------+---------+---------+---------+---------+---------  50799
        CCACGCCACCCGCCGCCCGCCGGCGAAGGCAAGTCGCCACTTGTCCGTGGCCTTGACACG
   31     G  A  V  G  G  G  R  P  L  P  F  S  G  E  Q  A  P  E  L  C    -

        GGCGTCGTGGTCACGTTCCACTACATCAACCGCATGGTCTCCCTCTTCCTCGACGACTCC
  50800 +---------+---------+---------+---------+---------+---------  50859
        CCGCAGCACCAGTGCAAGGTGATGTAGTTGGCGTACCAGAGGGAGAAGGAGCTGCTGAGG
   31     G  V  V  V  T  F  H  Y  I  N  R  M  V  S  L  F  L  D  D  S    -

        CCCATGCCGACCCGGACGCCGACACCGTTGCGCGGGCCCATCATGAGGACCACCGCACTG
  50860 +---------+---------+---------+---------+---------+---------  50919
        GGGTACGGCTGGGCCTGCGGCTGTGGCAACGCGCCCGGGTAGTACTCCTGGTGGCGTGAC
   31     P  M  P  T  R  T  P  T  P  L  R  G  P  I  M  R  T  T  A  L    -

        GCCATGCGTCCCGTCGGCCCGGGGCTGCTGACACCGGGCGCATCGCTCGGCCTGCTGCCT
  50920 +---------+---------+---------+---------+---------+---------  50979
        CGGTACGCAGGGCAGCCGGGCCCCGACGACTGTGGCCCGCGTAGCGAGCCGGACGACGGA
   31     A  M  R  P  V  G  P  G  L  L  T  P  G  A  S  L  G  L  L  P    -

        CCGGCTCCCCTGCCGCCCGGACTGGAGTGGGCCGAGGGCAACCCTTTCGTGGCCCAGGCG
  50980 +---------+---------+---------+---------+---------+---------  51039
        GGCCGAGGGGACGGCGGGCCTGACCTCACCCGGCTCCCGTTGGGAAAGCACCGGGTCCGC
   31     P  A  P  L  P  P  G  L  E  W  A  E  G  N  P  F  V  A  Q  A    -

        CTGGGGCGTGCCGTCGCCGCTGTGGACCAGGGAGCGCACTGGGTGCCCGAACCGGTCCGG
  51040 +---------+---------+---------+---------+---------+---------  51099
        GACCCCGCACGGCAGCGGCGACACCTGGTCCCTCGCGTGACCCACGGGCTTGGCCAGGCC
   31     L  G  R  A  V  A  A  V  D  Q  G  A  H  W  V  P  E  P  V  R    -

        GAGCGGCTGCGCACACGTCTGGACACCTGGGACGGATCGGCGCCGGGCCTCGGCCGGGGA
  51100 +---------+---------+---------+---------+---------+---------  51159
        CTCGCCGACGCGTGTGCAGACCTGTGGACCCTGCCTAGCCGCGGCCCGGAGCCGGCCCCT
   31     E  R  L  R  T  R  L  D  T  W  D  G  S  A  P  G  L  G  R  G    -

        TGGCTCGACGAGGCCGTGTCCGGCCTGCCGCCCCCAGGACGTGCCCGCGGCACGGCTGGCG
  51160 +---------+---------+---------+---------+---------+---------  51219
        ACCGAGCTGCTCCGGCACAGGCCGGACGGCGGGGTCCTGCACGGGCGCCGTGCCGACCGC
   31     W  L  D  E  A  V  S  G  L  P  P  Q  D  V  P  A  A  R  L  A    -

        CTGCTGACGGCCTTCGCCCCCCTACCAGGTGCTCCCGGACGACGTCGAGGAGTTCAGACGG
```

64

```
                                    ●                              ●



51220 +---------+---------+---------+---------+---------+--------- 51279
      GACGACTGCCGGAAGCGGGGGATGGTCCACGAGGGCCTGCTGCAGCTCCTCAAGTCTGCC
   31 GCAGCCGGGTGGCTGGCGCTTGAGCAGCTCGAGTGCAGGATGCGGCGCGACTGGTGCCGG
         L  L  T  A  F  A  P  Y  Q  V  L  P  D  D  V  E  E  F  R  R   -


      CGTCGGCCCACCGACCGCGAACTCGTCGAGCTCACGTCCTACGCCGCGCTGACCACGGCC
51280 +---------+---------+---------+---------+---------+--------- 51339
   31 GCAGCCGGGTGGCTGGCGCTTGAGCAGCTCGAGTGCAGGATGCGGCGCGACTGGTGCCGG
         R  R  P  T  D  R  E  L  V  E  L  T  S  Y  A  A  L  T  T  A   -


      GTCCGTGTCGGTCGCACGCTCGTCGTGCCCGACGCCGCCGGGCCGGGATGAACGGCCCCG
51340 +---------+---------+---------+---------+---------+--------- 51399
   31 CAGGCACAGCCAGCGTGCGAGCAGCACGGGCTGCGGCGGCCCGGCCCTACTTGCCGGGGC
         V  R  V  G  R  T  L  V  V  P  D  A  A  G  P  G  *  (ORF31)


      CAACGGCTCGGGAAGGCTGTCTCACGGCCGGAGGCGTACGCCGGTGAGGTGCTCGGACTC
51400 +---------+---------+---------+---------+---------+--------- 51459
      GTTGCCGAGCCCTTCCGACAGAGTGCCGGCCTCCGCATGCGGCCACTCCACGAGCCTGAG
           (ORF32)    *  P  R  L  R  V  G  T  L  H  E  S  E   -


      CTCCCAGAGGCGGCGCCGGGCCCTGGGGTCGACGGCTGCTCCGCCGGGGCGCACGAGCCC
51460 +---------+---------+---------+---------+---------+--------- 51519
   32 GAGGGTCTCCGCCGCGGCCCGGGACCCCAGCTGCCGACGAGGCGGCCCCGCGTGCTCGGG
         E  W  L  R  R  R  A  R  P  D  V  A  A  G  G  P  R  V  L  G   -


      GGGTGCGCCCCGGGTCTCGGTCACGCCGAGGGGCCCGTAGAACTCGCCCCCGCGCGCGCC
51520 +---------+---------+---------+---------+---------+--------- 51579
   32 CCCACGCGGGGGCCCAGAGCCAGTGCGGCTCCCCGGGCATCTTGAGCGGGGGCGCGCGCGG
         P  A  G  R  T  E  T  V  G  L  P  G  Y  F  E  G  G  R  A  G   -


      GGGATCGGTGGCCGCCCGCAGACCAGGCAGCATCCCCGCCGCGGCGGGCTGCAGGAACAA
51580 +---------+---------+---------+---------+---------+--------- 51639
   32 CCCTAGCCACCGGCGGGCGTCTGGTCCGTCGTAGGGGCGGCGCCGCCCGACGTCCTTGTT
         P  D  T  A  A  R  L  G  P  L  M  G  A  A  A  P  Q  L  F  L   -


      CGGGGCGAGCGGGGAGCCGAGCCTGCGCACGGGCGCGGGAAAGTCCCGGCCCAGACCGGT
51640 +---------+---------+---------+---------+---------+--------- 51699
   32 GCCCCGCTCGCCCCTCGGCTCGGACGCGTGCCCGCGCCCTTTCAGGGCCGGGTCTGGCCA
         P  A  L  P  S  G  L  R  R  V  P  A  P  F  D  R  G  L  G  T   -


      CGCGGTCAGCCCGGGATGAGCGGCGAGCGAGGCCAGTTCCGCGCCGGACTCCGCCAGTCT
51700 +---------+---------+---------+---------+---------+--------- 51759
   32 GCGCCAGTCGGGCCCTACTCGCCGCTCGCTCCGGTCAAGGCGCGGCCTGAGGCGGTCAGA
         A  T  L  G  P  H  A  A  L  S  A  L  E  A  G  S  E  A  L  R   -


      GTGATGGAGTTCCAGCGCGAACATGAGGTTGGCCAGCTTGGACTGGTTGTAGGCCCGGTA
51760 +---------+---------+---------+---------+---------+--------- 51819
   32 CACTACCTCAAGGTCGCGCTTGTACTCCAACCGGTCGAACCTGACCAACATCCGGGCCAT
          H  H  L  E  L  A  F  M  L  N  A  L  K  S  Q  N  Y  A  R  Y   -


      CCGGCTGTAGCGGCGTTCGCCGTGAAGGTCGCTGAAGTCGATGCGCCCCAGCCGGTGCAG
51820 +---------+---------+---------+---------+---------+--------- 51879
   32 GGCCGACATCGCCGCAAGCGGCACTTCCAGCGACTTCAGCTACGCGGGGTCGGCCACGTC
         R  S  Y  R  R  E  G  H  L  D  S  F  D  I  R  G  L  R  H  L   -


      ATAGCTGCTGATCGTCACGACCCGCGCGCCCGGCGCGGCCCGCAGGCTGTCCAGGAGCAG
51880 +---------+---------+---------+---------+---------+--------- 51939
   32 TATCGACGACTAGCAGTGCTGGGCGCGCGGGCCGCGCCGGGCGTCCGACAGGTCCTCGTC
         Y  S  S  I  T  V  V  R  A  G  P  A  A  R  L  S  D  L  L  L   -


      GCCGGTGAGGGCGAAGTGCCCCAGGTGGTTCGTGGCGAACTGGAGTTCGTGACCGTCCGG
51940 +---------+---------+---------+---------+---------+--------- 51999
   32 CGGCCACTCCCGCTTCACGGGGTCCACCAAGCACCGCTTGACCTCAAGCACTGGCAGGCC
         G  T  L  A  F  H  G  L  H  N  T  A  F  Q  L  E  H  G  D  P   -


      GGTGCGGGCCCGGTCGGTCCACATCACGCCCGCGTTGTTGACCAGCAGGTGGATGCGCGG
52000 +---------+---------+---------+---------+---------+--------- 52059
```

65

```
                   CCACGCCCGGGCCAGCCAGGTGTAGTGCGGGCGCAACAACTGGTCGTCCACCTACGCGCC
       32              T  R  A  R  D  T  W  M  V  G  A  N  N  V  L  L  H  I  R  P -

                   GAAGCGGTCGCGCAGTTCCTCGGCGCCGGCACGCACCGACGCGAGACGGGAAAGATCCAG
       52060       +---------+---------+---------+---------+---------+--------- 52119
                   CTTCGCCAGCGCGTCAAGGAGCCGCGGCCGTGCGTGGCTGCGCTCTGCCCTTTCTAGGTC
       32              F  R  D  R  L  E  E  A  G  A  R  V  S  A  L  R  S  L  D  L -

                   CCGTCTGACCGTCAGTTGCGCCGACGGCACCGGCTTTGGATGCGGGCCGCCGCGGCGAC
       52120       +---------+---------+---------+---------+---------+--------- 52179
                   GGCAGACTGGCAGTCAACGCGGCTGCCGTGGGCCGAAACCTACGCCCGGCGGCGCCGCTG
       32              R  R  V  T  L  Q  A  S  P  V  R  S  Q  I  R  A  A  A  A  V -

                   CCCGCGGTCCGGATCGCGCACGGCCAGCACCACGTGGGCGCCGTGCCGGGCGAGCTCCTG
       52180       +---------+---------+---------+---------+---------+--------- 52239
                   GGGCGCCAGGCCTAGCGCGTGCCGGTCGTGGTGCACCCGCGGCACGGCCCGCTCGAGGAC
       32              G  R  D  P  D  R  V  A  L  V  V  H  A  G  H  R  A  L  E  Q -

                   CGCCAGGTGCAGTCCGATGCCGGAGCTGGCACCGGTGACCACCGCGGTGGTTCCGGTACG
       52240       +---------+---------+---------+---------+---------+--------- 52299
                   GCGGTCCACGTCAGGCTACGGCCTCGACCGTGGCCACTGGTGGCGCCACCAAGGCCATGC
       32              A  L  H  L  G  I  G  S  S  A  G  T  V  V  A  T  T  G  T  R -

                   GTCCGGGACATCGGCGGCGCTCCAGCGTCGCCGCGTTCTCATCGGTCGTCCCTCCCGGGG
       52300       +---------+---------+---------+---------+---------+--------- 52359
                   CAGGCCCTGTAGCCGCCGCGAGGTCGCAGCGGCGCAAGAGTAGCCAGCAGGGAGGGCCCC
       32              D  P  V  D  A  A  S  W  R  R  R  T  R  M  (ORF32)

                   GATGCGTCAGCCGGCCTGGGCCATCGCGGCCCGGTAGCCGTTGGCGACGATCTGCCGGGC
       52360       +---------+---------+---------+---------+---------+--------- 52419
                   CTACGCAGTCGGCCGGACCCGGTAGCGCCGGGCCATCGGCAACCGCTGCTAGACGGCCCG

                   GGAGTGCTCGTAGTACTCGTCGTCCTTCGGCAGCTCCGTGGCGAGACCGCTGACGTACCG
       52420       +---------+---------+---------+---------+---------+--------- 52479
                   CCTCACGAGCATCATGAGCAGCAGGAAGCCGTCGAGGCACCGCTCTGGCGACTGCATGGC

                   GTTGAACATGCAGAACGCGGCGGCGATCAGAACGGTGTCGTGCAGAGCGGTGTCGTCCGC
       52480       +---------+---------+---------+---------+---------+--------- 52539
                   CAACTTGTACGTCTTGCGCCGCCGCTAGTCTTGCCACAGCACGTCTCGCCACAGCAGGCG

                   TCCCTCGGCCCGCGCCGAGGCGATCACCCCTGCGGAGACCGGGCGCGCCGCGCTCTGGAC
       52540       +---------+---------+---------+---------+---------+--------- 52599
                   AGGGAGCCGGGCGCGGCTCCGCTAGTGGGGACGCCTCTGGCCCGCGCGGCGCGAGACCTG

                   CTCGGCGGCGACGGCCAGCAGCGCGCGCGTCCTGCCGTCGATGGGCGCGGTGGCGGGGTC
       52600       +---------+---------+---------+---------+---------+--------- 52659
                   GAGCCGCCGCTGCCGGTCGTCGCGCGCGCAGGACGGCAGCTACCCGCGCCACCGCCCCAG

                   GGCGAGGACGGCCTCGACGAGCTGCCGGCCTCCCGGCAGCTGCGCGGCGGCGAAGGCCCC
       52660       +---------+---------+---------+---------+---------+--------- 52719
                   CCGCTCCTGCCGGAGCTGCTCGACGGCCGGAGGGCCGTCGACGCGCCGCCGCTTCCGGGG

                   GTGGGAGGCGGCGCAGAACTCGGTGGAGTTGAGATGCGAGACGTACGCCGCGATGAGCTC
       52720       +---------+---------+---------+---------+---------+--------- 52779
                   CACCCTCCGCCGCGTCTTGAGCCACCTCAACTCTACGCTCTGCATGCGGCGCTACTCGAG

                   GCGTTGCCCCGGTTCCAGCGAGGACGGCGCCCGCAGCAGGGCGTTCGCGAGATCGCCCAG
       52780       +---------+---------+---------+---------+---------+--------- 52839
                   CGCAACGGGGCCAAGGTCGCTCCTGCCGCGGGCGTCGTCCCGCAAGCGCTCTAGCGGGTC

                   CGGTGCTGCGGTGCCGGGGTGGTGAGCCATCAGACCACTGATGCCGGGGAGGTCGTTGTC
       52840       +---------+---------+---------+---------+---------+--------- 52899
                   GCCACGACGCCACGGCCCCACCACTCGGTAGTCTGGTGACTACGGCCCCTCCAGCAACAG

                   GAGTGCTATGTGGGGCACGGCTCTTCCTTCCGGGTGGACGAGGGGCGGACGGCGGCGGAT
       52900       +---------+---------+---------+---------+---------+--------- 52959
```

66

```
              CTCACGATACACCCCGTGCCGAGAAGGAAGGCCCACCTGCTCCCCGCCTGCCGCCGCCTA
              CAGGGCCATTCGACTTCGTCGTCGGCGGCCGCGCAGATGCGGGTGAAGGGCCATTCCACG
      52960   +---------+---------+---------+---------+---------+---------  53019
              GTCCCGGTAAGCTGAAGCAGCAGCCGCCGGCGCGTCTACGCCCACTTCCCGGTAAGGTGC

              TCTTCCCCTCCCGTTGCGGAGTGGGCGGAGGCCGTGGTGAAGAGGGTGACGAGTCCGAAC
      53020   +---------+---------+---------+---------+---------+---------  53079
              AGAAGGGGAGGGCAACGCCTCACCCGCCTCCGGCACCACTTCTCCCACTGCTCAGGCTTG

              GTGCCGAAGAGGAGGGACAGTCGGGCAACGTGAAGTGCGGTACCCATGCGAGCTCCTAGC
      53080   +---------+---------+---------+---------+---------+---------  53139
              CACGGCTTCTCCTCCCTGTCAGCCCGTTGCACTTCACGCCATGGGTACGCTCGAGGATCG

              GAGGGCGGCGTGACCGCGGGACGGTGAGACCTCGTGATGCCAGGAAGCTAGCGAATCGGA
      53140   +---------+---------+---------+---------+---------+---------  53199
              CTCCCGCCGCACTGGCGCCCTGCCACTCTGGAGCACTACGGTCCTTCGATCGCTTAGCCT

              CTGAGGGTGGCAACGATATGCCAGACTTTGGCAACTTGCCTGTGTATCAGCCGGACTGTC
      53200   +---------+---------+---------+---------+---------+---------  53259
              GACTCCCACCGTTGCTATACGGTCTGAAACCGTTGAACGGACACATAGTCGGCCTGACAG
    33                                                  (ORF33) V  Y  Q  P  D  C  R

              GGCCGCTGGTAAAGACGGAACGGCGAGATCCCGCGACCGCGTCGCAGAGCAGCAGGGTCT
      53260   +---------+---------+---------+---------+---------+---------  53319
              CCGGCGACCATTTCTGCCTTGCCGCTCTAGGGCGCTGGCGCAGCGTCTCGTCGTCCCAGA
    33           P  L  V  K  T  E  R  R  D  P  A  T  A  S  Q  S  S  R  V  C -

              GCTCACCCAGCGTCGGGGCGGCCAGCATGTCGCGTACCGGGAGCGTGACGCCCAGCTCGC
      53320   +---------+---------+---------+---------+---------+---------  53379
              CGAGTGGGTCGCAGCCCCGCCGGTCGTACAGCGCATGGCCCTCGCACTGCGGGTCGAGCG
    33           S  P  S  V  G  A  A  S  M  S  R  T  G  S  V  T  P  S  S  R -

              GGTTGATCCTGCGGACCAGCCGGGTGATGAGCAGGGAGTCGCCGCCGTGGGCGAAGAAAT
      53380   +---------+---------+---------+---------+---------+---------  53439
              CCAACTAGGACGCCTGGTCGGCCCACTACTCGTCCCTCAGCGGCGGCACCCGCTTCTTTA
    33           L  I  L  R  T  S  R  V  M  S  R  E  S  P  P  W  A  K  K  S -

              CAGCACCTTCGGAGGGGTCCGGGAAGCCGAGCAGGTCACCCCAGCCGCGCACCAGTACCT
      53440   +---------+---------+---------+---------+---------+---------  53499
              GTCGTGGAAGCCTCCCCAGGCCCTTCGGCTCGTCCAGTGGGGTCGGCGCGTGGTCATGGA
    33           A  P  S  E  G  S  G  K  P  S  R  S  P  Q  P  R  T  S  T  W -

              GGCGGATGTCGCCGGTGGTGACGACCGTGCGCCGGGAGCCCCGACGTGCCGAGCGCAGCC
      53500   +---------+---------+---------+---------+---------+---------  53559
              CCGCCTACAGCGGCCACCACTGCTGGCACGCGGCCCTCGGGGCTGCACGGCTCGCGTCGG
    33           R  M  S  P  V  V  T  T  V  R  R  E  P  R  R  A  E  R  S  R -

              GCGAGGCATGCACCAGCGCCACCTGGTCGCCGAGGTTGCGCCGCGACAGCTCGCGCAGCG
      53560   +---------+---------+---------+---------+---------+---------  53619
              CGCTCCGTACGTGGTCGCGGTGGACCAGCGGCTCCAACGCGGCGCTGTCGAGCGCGTCGC
    33           E  A  C  T  S  A  T  W  S  P  R  L  R  R  D  S  S  R  S  D -

              ACACCGTGACGCCGAACCTCTCGGTGATCCTGCGGACCAGCCGCGTGATCAGCAGCGTGT
      53620   +---------+---------+---------+---------+---------+---------  53679
              TGTGGCACTGCGGCTTGGAGAGCCACTAGGACGCCTGGTCGGCGCACTAGTCGTCGCACA
    33           T  V  T  P  N  L  S  V  I  L  R  T  S  R  V  I  S  S  V  S -

              CCCCGCCGCGCGCGAAGAAATCCGAATGCTCGGTGAGGTCGGAGCGGCCGAGGAGCTCGC
      53680   +---------+---------+---------+---------+---------+---------  53739
              GGGGCGGCGCGCGCTTCTTTAGGCTTACGAGCCACTCCAGCCTCGCCGGCTCCTCGAGCG
    33           P  P  R  A  K  K  S  E  C  S  V  R  S  E  R  P  R  S  S  L -

              TCCACGCGCCGACCATGAACTCCCCCACGTCACCGAGCCGGTGCTCGTCGCCGTCGGGGC
      53740   +---------+---------+---------+---------+---------+---------  53799
              AGGTGCGCGGCTGGTACTTGAGGGGGTGCAGTGGCTCGGCCACGAGCAGCGGCAGCCCCG
```

67

```
33              H  A  P  T  M  N  S  P  T  S  P  S  R  C  S  S  P  S  G  P -
        CCTTCGGCGCGCCGGATCCCGCGGAACGGTTCCGGCCGGAGACGGCAGAGCGGTCACTGG
  53800 +---------+---------+---------+---------+---------+---------  53859
        GGAAGCCGCGCGGCCTAGGGCGCCTTGCCAAGGCCGGCCTCTGCCGTCTCGCCAGTGACC
  33              F  G  A  P  D  P  A  E  R  F  R  P  E  T  A  E  R  S  L  V -

        TCACTTTCGCCACCTCCAGGGGCATGTGTCGGCTGCATCGGCTTCCCGCCACGGTACGGG
  53860 +---------+---------+---------+---------+---------+---------  53919
        AGTGAAAGCGGTGGAGGTCCCCGTACACAGCCGACGTAGCCGAAGGGCGGTGCCATGCCC
  33              T  F  A  T  S  R  G  M  C  R  L  H  R  L  P  A  T  V  R  E -

        AGCACATGTTGCATGGCAATACCTTTCCAAGTCGGTGGCAACCCTCCTTGCCATCCACCC
  53920 +---------+---------+---------+---------+---------+---------  53979
        TCGTGTACAACGTACCGTTATGGAAAGGTTCAGCCACCGTTGGGAGGAACGGTAGGTGGG
  33              H  M  L  H  G  N  T  F  P  S  R  W  Q  P  S  L  P  S  T  H -

        ACTGCAGTTGGGCGAGATGTGTAGGCATTCGAGGTCCGCAGGTTTGCCAAGCCGCGCGCG
  53980 +---------+---------+---------+---------+---------+---------  54039
        TGACGTCAACCCGCTCTACACATCCGTAAGCTCCAGGCGTCCAAACGGTTCGGCGCGCGC
  33               C  S  W  A  R  C  V  G  I  R  G  P  Q  V  C  Q  A  A  R  D -

        ACCGGCATACTCTCTGGCACAACTGGAATGAGTAGCGTGGCAGGCCACGGGGACCGGGCC
  54040 +---------+---------+---------+---------+---------+---------  54099
        TGGCCGTATGAGAGACCGTGTTGACCTTACTCATCGCACCGTCCGGTGCCCCTGGCCCGG
  33               R  H  T  L  W  H  N  W  N  E  *  (ORF33)

        GGGCCAGGAACCTTCGTCCTCCATCTATTCGCTGGGGCGTGCACGTGTTGGAGCAGCCAT
  54100 +---------+---------+---------+---------+---------+---------  54159
        CCCGGTCCTTGGAAGCAGGAGGTAGATAAGCGACCCCGCACGTGCACAACCTCGTCGGTA

        CTTTCGGCCGTCGCCTGAGGCAGCTGAGGACCGAGCGGGGTCTTTCCCAGGCCGCGCTCG
  54160 +---------+---------+---------+---------+---------+---------  54219
        GAAAGCCGGCAGCGGACTCCGTCGACTCCTGGCTCGCCCCAGAAAGGGTCCGGCGCGAGC

        CGGGGGACGGCATGTCTACGGGCTATCTCTCGCGCCTGGAGTCGGGCGCCCGGCAGCCCT
  54220 +---------+---------+---------+---------+---------+---------  54279
        GCCCCCTGCCGTACAGATGCCCGATAGAGAGCGCGGACCTCAGCCCGCGGGCCGTCGGGA
  34              (ORF34) M  S  T  G  Y  L  S  R  L  E  S  G  A  R  Q  P  S -

        CCGATCGCGCCGTCGCCCACCTGGCCGGACAACTCGGCATCAGCCCGTCGGAGTTCGAAG
  54280 +---------+---------+---------+---------+---------+---------  54339
        GGCTAGCGCGGCAGCGGGTGGACCGGCCTGTTGAGCCGTAGTCGGGCAGCCTCAAGCTTC
  34              D  R  A  V  A  H  L  A  G  Q  L  G  I  S  P  S  E  F  E  G -

        GGTCCCGGGCCACCTCGCTCGCCCAGATCCTCTCCCTCTCCACTTCCCTGGAGTCCGACG
  54340 +---------+---------+---------+---------+---------+---------  54399
        CCAGGGCCCGGTGGAGCGAGCGGGTCTAGGAGAGGGAGAGGTGAAGGGACCTCAGGCTGC
  34              S  R  A  T  S  L  A  Q  I  L  S  L  S  T  S  L  E  S  D  E -

        AGACCAGTGAGCTTCTCGCCGAGGCGGTACGTTCCGCGCATGGCCAGGATCCGATGCTCC
  54400 +---------+---------+---------+---------+---------+---------  54459
        TCTGGTCACTCGAAGAGCGGCTCCGCCATGCAAGGCGCGTACCGGTCCTAGGCTACGAGG
  34              T  S  E  L  L  A  E  A  V  R  S  A  H  G  Q  D  P  M  L  R -

        GCTGGCAGGCCCTGTGGCTGCTGGGACAGTGGAAGCGCCGGCACGGCGACTCGGCCGGCG
  54460 +---------+---------+---------+---------+---------+---------  54519
        CGACCGTCCGGGACACCGACGACCCTGTCACCTTCGCGGCCGTGCCGCTGAGCCGGCCGC
  34              W  Q  A  L  W  L  L  G  Q  W  K  R  R  H  G  D  S  A  G  E -

        AGCACGGCTACCTCCAGCGTCTGGTGACGCTGAGTGAGGAGATCGGCCTGGCCGAGTTGC
  54520 +---------+---------+---------+---------+---------+---------  54579
        TCGTGCCGATGGAGGTCGCAGACCACTGCGACTCACTCCTCTAGCCGGACCGGCTCAACG
  34              H  G  Y  L  Q  R  L  V  T  L  S  E  E  I  G  L  A  E  L  R -

        GCGCACGGGCCCTGACCCAGTTCGCCCGGTCGCTGCGGGTACTGGGCGAGATCGTTCCGG
```

68

```
              ●                                    ●

      CGCGTGCCCGGGACTGGGTCAAGCGGGCCAGCGACGCCCATGACCCGCTCTAGCAAGGCC
54580 +---------+---------+---------+---------+---------+--------- 54639
34         A  R  A  L  T  Q  F  A  R  S  L  R  V  L  G  E  I  V  P  A -

      CGGTGGAGGCTGCCGCCGCCGCCCACCGGCTCGCGGTGGACCATGCGCTGTCCAGCCAGG
54640 +---------+---------+---------+---------+---------+--------- 54699
      GCCACCTCCGACGGCGGCGGCGGGTGGCCGAGCGCCACCTGGTACGCGACAGGTCGGTCC
34         V  E  A  A  A  A  A  H  R  L  A  V  D  H  A  L  S  S  Q  D -

      ACAGGGCCGCTTCGCTGCTGGTTCTGGTGTCGGTGGAGGCCGAGGCGGGACGGATGCCCG
54700 +---------+---------+---------+---------+---------+--------- 54759
      TGTCCCGGCGAAGCGACGACCAAGACCACAGCCACCTCCGGCTCCGCCCTGCCTACGGGC
34         R  A  A  S  L  L  V  L  V  S  V  E  A  A  G  R  M  P  D -

      ACGCCCGGCGCCACGCCGACGAACTGACCGTCCTGGTGAGGGGACGGTCCGACACTCTGT
54760 +---------+---------+---------+---------+---------+--------- 54819
      TGCGGGCCGCGGTGCGGCTGCTTGACTGGCAGGACCACTCCCCTGCCAGGCTGTGAGACA
34         A  R  R  H  A  D  E  L  T  V  L  V  R  G  R  S  D  T  L  W -

      GGGCCGAGGCGTTGTGGACGGCGGGTGCGTTGAAGGTGCGGCAGGGCGAGTTCGCCGCGG
54820 +---------+---------+---------+---------+---------+--------- 54879
      CCCGGCTCCGCAACACCTGCCGCCCACGCAACTTCCACGCCGTCCCGCTCAAGCGGCGCC
34         A  E  A  L  W  T  A  G  A  L  K  V  R  Q  G  E  F  A  A  A -

      CCGAGGTCCTTTTCCAGGAGGCTCTGGACGGGTTCGACAGCCGGGAGAACCTGACGATCT
54880 +---------+---------+---------+---------+---------+--------- 54939
      GGCTCCAGGAAAAGGTCCTCCGAGACCTGCCCAAGCTGTCGGCCCTCTTGGACTGCTAGA
34         E  V  L  F  Q  E  A  L  D  G  F  D  S  R  E  N  L  T  I  W -

      GGCTGCGGCTGCGCATCGCGATGGCCGAACTCCACCTGCAGAAACTTCCTCCCGAGCCCG
54940 +---------+---------+---------+---------+---------+--------- 54999
      CCGACGCCGACGCGTAGCGCTACCGGCTTGAGGTGGACGTCTTTGAAGGAGGGCTCGGGC
34         L  R  L  R  I  A  M  A  E  L  H  L  Q  K  L  P  P  E  P  D -

      ACGCCGCGCAGCTCTGCATCGAGGCGGCGGAGGCGGCCCTTCCCTTTGCCCGCACATCCG
55000 +---------+---------+---------+---------+---------+--------- 55059
      TGCGGCGCGTCGAGACGTAGCTCCGCCGCCTCCGCCGGGAAGGGAAACGGGCGTGTAGGC
34         A  A  Q  L  C  I  E  A  A  E  A  A  L  P  F  A  R  T  S  A -

      CTCTGGAACAGTCCCTCGCCGCTCTGCGGGCGCGCCTCGCCTTCCATGAGGGCAGGTTCG
55060 +---------+---------+---------+---------+---------+--------- 55119
      GAGACCTTGTCAGGGAGCGGCGAGACGCCCGCGCGGAGCGGAAGGTACTCCCGTCCAAGC
34         L  E  Q  S  L  A  A  L  R  A  R  L  A  F  H  E  G  R  F  A -

      CCGATGCCCGCGCGTTGTTGGAGAGGCTCGGCAGGACCGAGCTCCGGCTGCCCTATCAGA
55120 +---------+---------+---------+---------+---------+--------- 55179
      GGCTACGGGCGCGCAACAACCTCTCCGAGCCGTCCTGGCTCGAGGCCGACGGGATAGTCT
34         D  A  R  A  L  L  E  R  L  G  R  T  E  L  R  L  P  Y  Q  S -

      GCCGGATCCGCCTGGAGGTCCTCGGTCATCAGCTGCGCATCCTGAGCGGGGAGGAGGAGG
55180 +---------+---------+---------+---------+---------+--------- 55239
      CGGCCTAGGCGGACCTCCAGGAGCCAGTAGTCGACGCGTAGGACTCGCCCCTCCTCCTCC
34         R  I  R  L  E  V  L  G  H  Q  L  R  I  L  S  G  E  E  E -

      AAGGCCTGGCCGGCCTCCAGCTCCTGGCCGAGGAGGCGCAGGAGAACTCCAACATCAACC
55240 +---------+---------+---------+---------+---------+--------- 55299
      TTCCGGACCGGCCGGAGGTCGAGGACCGGCTCCTCCGCGTCCTCTTGAGGTTGTAGTTGG
34         G  L  A  G  L  Q  L  L  A  E  E  A  Q  E  N  S  N  I  N  L -

      TCGCCGCGGAGATCTGGCGGCTCGCGGCGGAATGCCTGATGCGGGCGCGCGGGAAGGTCC
55300 +---------+---------+---------+---------+---------+--------- 55359
      AGCGGCGCCTCTAGACCGCCGAGCGCCGCCTTACGGACTACGCCCGCGCGCCCTTCCAGG
34         A  A  E  I  W  R  L  A  A  E  C  L  M  R  A  R  G  K  V  R -

      GCGGCGCCACCGGCGGCTGACGCCGCGCCGGTTCGCGAGGTCCACCGCGCCGCCGTGGCC
55360 +---------+---------+---------+---------+---------+--------- 55419
```

```
                   CGCCGCGGTGGCCGCCGACTGCGGCGCGGCCAAGCGCTCCAGGTGGCGCGGCGGCACCGG
34                     G   A   T   G   G   *  (ORF34)

                   ACCGCCGTCGGCGTGAGGCGCCGGCGTGTGCCGCCCCCCACGGTTGCTCGCCCTTGGTGG
           55420   +---------+---------+---------+---------+---------+---------   55479
                   TGGCGGCAGCCGCACTCCGCGGCCGCACACGGCGGGGGGTGCCAACGAGCGGGAACCACC

                   TGCATCTGTTGGCACATGTGTACCTCCTACACAGTCAATTGTTGCCAAAATTGTCGAACC
           55480   +---------+---------+---------+---------+---------+---------   55539
                   ACGTAGACAACCGTGTACACATGGAGGATGTGTCAGTTAACAACGGTTTTAACAGCTTGG

                   GAATGGCAATTGCTTGCCTTTGCTGAAGAGGCGTGCTGATATGCAAGTCAAGTAGCCTCC
           55540   +---------+---------+---------+---------+---------+---------   55599
                   CTTACCGTTAACGAACGGAAACGACTTCTCCGCACGACTATACGTTCAGTTCATCGGAGG

                   TCCGATCTCGGGCGGCCATATGGGAAACATCGAGTTGAGCGGCGATGGCGTTCGTCAGTG
           55600   +---------+---------+---------+---------+---------+---------   55659
                   AGGCTAGAGCCCGCCGGTATACCCTTTGTAGCTCAACTCGCCGCTACCGCAAGCAGTCAC

                   CTGCCGTTCTGGCCAGGCAACTGATGTCGATGGGGATGGCAAGATTTTGCCGAAAACCGA
           55660   +---------+---------+---------+---------~+---------+---------   55719
                   GACGGCAAGACCGGTCCGTTGACTACAGCTACCCCTACCGTTCTAAAACGGCTTTTGGCT

                   TACATCTCTGTCCGTCCCGGACAGCCTTCGCCCCCCGGGTGACACTGCTCCGGCATGGCT
           55720   +---------+---------+---------+---------+---------+---------   55779
                   ATGTAGAGACAGGCAGGGCCTGTCGGAAGCGGGGGGCCCACTGTGACGAGGCCGTACCGA

                   CCGGTTTCTCGTCGCCCGGCCGACGGACCGCACCGTCCGGAACGAGGCGCCGGTGTGCGT
           55780   +---------+---------+---------+---------+---------+---------   55839
                   GGCCAAAGAGCAGCGGGCCGGCTGCCTGGCGTGGCAGGCCTTGCTCCGCGGCCACACGCA

                   CCGCTGATGGGCACAGCGGCCTCGGCCGCAGCAGGTTCCCACCGAGAAGAATGCCGAGGC
           55840   +---------+---------+---------+---------+---------+---------   55899
                   GGCGACTACCCGTGTCGCCGGAGCCGGCGTCGTCCAAGGGTGGCTCTTCTTACGGCTCCG

                   CCAGCCGTGAACCACGACATGTCCCAGCGTGCCTTGCTGGAGGCGGCGGCCGAGGGGCTG
           55900   +---------+---------+---------+---------+---------+---------   55959
                   GGTCGGCACTTGGTGCTGTACAGGGTCGCACGGAACGACCTCCGCCGCCGGCTCCCCGAC

                   CGGCGGCTGGCCGGCGACGCGCGGTGCCGGAGCGCGTCGGCCGCGCCCTCCTCGGCATTG
           55960   +---------+---------+---------+---------+---------+--------~-   56019
                   GCCGCCGACCGGCCGCTGCGCGCCACGGCCTCGCGCAGCCGGCGCGGGAGGAGCCGTAAC

                   AGGGACATGTTCTCCCCCGCCGCCCGCCGGTACGTGCTCGCCTCGGACCGCGCGGGGTTC
           56020   +---------+---------+---------+---------+---------+---------   56079
                   TCCCTGTACAAGAGGGGGCGGCGGGCGGCCATGCACGAGCGGAGCCTGGCGCGCCCCAAG
35                 (ORF35) M   F   S   P   A   A   R   R   Y   V   L   A   S   D   R   A   G   F

                   TTCGAGCAGGCTGTCCGGCTGCGCTCCCGGGGGTACCGGGTGAGCGCGGAGTTCGTCGGC
           56080   +---------+---------+---------+---------+---------+---------   56139
                   AAGCTCGTCCGACAGGCCGACGCGAGGGCCCCCATGGCCCACTCGCGCCTCAAGCAGCCG
35                    F   E   Q   A   V   R   L   R   S   R   G   Y   R   V   S   A   E   F   V   G   -

                   CCCGATCAGGGAGCCACCGACGCCCTCCACGCGGAGCACGTGGTCGAAGAGCACCTGAGG
           56140   +---------+---------+---------+---------+---------+---------   56199
                   GGGCTAGTCCCTCGGTGGCTGCGGGAGGTGCGCCTCGTGCACCAGCTTCTCGTGGACTCC
35                    P   D   Q   G   A   T   D   A   L   H   A   E   H   V   V   E   E   H   L   R   -

                   CTGCTCGATCAGGAGCCGGCCCCTGACCGGATCGGTGTGGACGTCTCCCGGATCGGCCTC
           56200   +---------+---------+---------+---------+---------+---------   56259
                   GACGAGCTAGTCCTCGGCCGGGGACTGGCCTAGCCACACCTGCAGAGGGCCTAGCCGGAG
35                    L   L   D   Q   E   P   A   P   D   R   I   G   V   D   V   S   R   I   G   L   -

                   GCCCACTCGGCGCAGACTGCCCTGCGCAACACCGGGCGGCTGGCTGCCGCTGCGGCGCTC
           56260   +---------+---------+---------+---------+---------+---------   56319
                   CGGGTGAGCCGCGTCTGACGGGACGCGTTGTGGCCCGCCGACCGACGGCGACGCCGCGAG
```

70

```
 35                 A  H  S  A  Q  T  A  L  R  N  T  G  R  L  A  A  A  A  A  L   -
           CGCGGGAGCGAGGTCGTCCTGCTCATGGAGGGGTCCGAGGACATCGACACCGTGCTGGCC
  56320   +---------+---------+---------+---------+---------+---------         56379
           GCGCCCTCGCTCCAGCAGGACGAGTACCTCCCCAGGCTCCTGTAGCTGTGGCACGACCGG
 35                 R  G  S  E  V  V  L  L  M  E  G  S  E  D  I  D  T  V  L  A   -

           GTCCATGACGCCCTGGTGAACCGTTACGACAACGTGGGGATCACCCTTCAGGCGCACCTG
  56380   +---------+---------+---------+---------+---------+---------         56439
           CAGGTACTGCGGGACCACTTGGCAATGCTGTTGCACCCCTAGTGGGAAGTCCGCGTGGAC
 35                 V  H  D  A  L  V  N  R  Y  D  N  V  G  I  T  L  Q  A  H  L   -

           CACCGCACCGTGGACGACGCCATGGCGGTCGCGGGTCCTGGCCGCACCGTGCGGCTGGTC
  56440   +---------+---------+---------+---------+---------+---------         56499
           GTGGCGTGGCACCTGCTGCGGTACCGCCAGCGCCCAGGACCGGCGTGGCACGCCGACCAG
 35                 H  R  T  V  D  D  A  M  A  V  A  G  P  G  R  T  V  R  L  V   -

           ATGGGCTCCTCGGCCGAGCCTGCCGGCACCGCTCTGTCCCGGGGCCCCGCTCTGGAGGAC
  56500   +---------+---------+---------+---------+---------+---------         56559
           TACCCGAGGAGCCGGCTCGGACGGCCGTGGCGAGACAGGGCCCCGGGGCGAGACCTCCTG
 35                 M  G  S  S  A  E  P  A  G  T  A  L  S  R  G  P  A  L  E  D   -

           CGGTACCTTGACCTCGCGGAGCTTCTCGTGGACCGTGGCGTCCGGCTGAGTCTGGCCACT
  56560   +---------+---------+---------+---------+---------+---------         56619
           GCCATGGAACTGGAGCGCCTCGAAGAGCACCTGGCACCGCAGGCCGACTCAGACCGGTGA
 35                 R  Y  L  D  L  A  E  L  L  V  D  R  G  V  R  L  S  L  A  T   -

           CCGGACGCCGAGGTCCTGGCCGGGGCGCAGGAGCGTGGTCTGCTCGAACGCGTCCAGGAC
  56620   +---------+---------+---------+---------+---------+---------         56679
           GGCCTGCGGCTCCAGGACCGGCCCCGCGTCCTCGCACCAGACGAGCTTGCGCAGGTCCTG
 35                 P  D  A  E  V  L  A  G  A  Q  E  R  G  L  L  E  R  V  Q  D   -

           ATCGAGATGCTCTACGGTGTGCGGCCCGAGCTGCTGCGCCGCCACCGGGCGGCGGGCCGC
  56680   +---------+---------+---------+---------+---------+---------         56739
           TAGCTCTACGAGATGCCACACGCCGGGCTCGACGACGCGGCGGTGGCCCGCCGCCCGGCG
 35                 I  E  M  L  Y  G  V  R  P  E  L  L  R  R  H  R  A  A  G  R   -

           CCCTGTCGCATCCACGCGGCCTACGGGATGAACTGGTGGCTTCCCCTGCTGCGGAGGCTG
  56740   +---------+---------+---------+---------+---------+---------         56799
           GGGACAGCGTAGGTGCGCCGGATGCCCTACTTGACCACCGAAGGGGACGACGCCTCCGAC
 35                 P  C  R  I  H  A  A  Y  G  M  N  W  W  L  P  L  L  R  R  L   -

           GCCGACAACCCGCCGATGGTGCTCAACGCCCTGGCCGACATCGGCCGGGACCGGGAGCCC
  56800   +---------+---------+---------+---------+---------+---------         56859
           CGGCTGTTGGGCGGCTACCACGAGTTGCGGGACCGGCTGTAGCCGGCCCTGGCCCTCGGG
 35                 A  D  N  P  P  M  V  L  N  A  L  A  D  I  G  R  D  R  E  P   -

           GTCGCCCACCAGGCGTACTGACCCGCCCCGGGCCGCGATCCGCGGGGCACCGGCCCCGGG
  56860   +---------+---------+---------+---------+---------+---------         56919
           CAGCGGGTGGTCCGCATGACTGGGCGGGGCCCGGCGCTAGGCGCCCCGTGGCCGGGGCCC
 35                 V  A  H  Q  A  Y  *  (OF35)

           GCGCCGGTCAGCTCCCGGTCGCCGCGAACTGCCCGGGCCTGCGCCCCTCGCCCGCCGGCC
  56920   +---------+---------+---------+---------+---------+---------         56979
           CGCGGCCAGTCGAGGGCCAGCGGCGCTTGACGGGCCCGGACGCGGGGAGCGGGCGGCCGG
 36                 (ORF36)  *  S  G  T  A  A  F  Q  G  P  R  R  G  E  G  A  P   -

           CCCGGTAGGCCTGGGCGATGTCCAGCCACTTCTCCGCCTCCTGACCAGACGCGGTCAGGG
  56980   +---------+---------+---------+---------+---------+---------         57039
           GGGCCATCCGGACCCGCTACAGGTCGGTGAAGAGGCGGAGGACTGGTCTGCGCCAGTCCC
 36                 G  R  Y  A  Q  A  I  D  L  W  K  E  A  E  Q  G  S  A  T  L   -

           CGAGGTCGTCGCGGTGGCGGCGCCGGGTGACCAGCAGGCAGAAGTCGTGCGCGGGACCGC
  57040   +---------+---------+---------+---------+---------+---------         57099
           GCTCCAGCAGCGCCACCGCCGCGGCCCACTGGTCGTCCGTCTTCAGCACGCGCCCTGGCG
 36                 A  L  D  D  R  H  R  R  R  T  V  L  L  C  F  D  H  A  P  G   -
```

71

```
          TGACCGTCTCGGTGGCGTCCTCGGGGCCGACCGTCCAGACCTCGCCCGAGGGGGCGGTGA
57100     +---------+---------+---------+---------+---------+---------   57159
          ACTGGCAGAGCCACCGCAGGAGCCCGGCTGGCAGGTCTGGAGCGGGCTCCCCCGCCACT
36          S  V  T  E  T  A  D  E  P  G  V  T  W  V  E  G  S  P  A  T   -

          GCTCGAAGCGGAACGGCGCGGCCGGCGGGGTCAGACCGTGGGACTCGTAGCCGAAGTCGC
57160     +---------+---------+---------+---------+---------+---------   57219
          CGAGCTTCGCCTTGCCGCGCCGGCCGCCCCAGTCTGGCACCCTGAGCATCGGCTTCAGCG
36          L  E  F  R  F  P  A  A  P  P  T  L  G  H  S  E  Y  G  F  D   -

          GTGTCAGCCAGGCGAAGTCGACGATGTTGCGAAGCCGCTCGGTGGGCGTGCGCCGGACAC
57220     +---------+---------+---------+---------+---------+---------   57279
          CACAGTCGGTCCGCTTCAGCTGCTACAACGCTTCGGCGAGCCACCCGCACGCGGCCTGTG
36          R  T  L  W  A  F  D  V  I  N  R  L  R  E  T  P  T  R  R  V   -

          CCAGGGCGTCGGCGACGTCCTGGCCGTGGGCGAACACCTCCATGATCCCGGCGCAGCCCA
57280     +---------+---------+---------+---------+---------+---------   57339
          GGTCCCGCAGCCGCTGCAGGACCGGCACCCGCTTGTGGAGGTACTAGGGCCGCGTCGGGT
36          G  L  A  D  A  V  D  Q  G  H  A  F  V  E  M  I  G  A  C  G   -

          GAACGACCGGCGGCAGCGGGTTGACCAGCCACGGAACCACCTGGCCGGCGGGGACCGCGG
57340     +---------+---------+---------+---------+---------+---------   57399
          CTTGCTGGCCGCCGTCGCCCAACTGGTCGGTGCCTTGGTGGACCGGCCGCCCCTGGCGCC
36          L  V  V  P  P  L  P  N  V  L  W  P  V  V  Q  G  A  P  V  A   -

          CGAGCGCCTCGACCGAGGCCCGCCCCATGCCCCGGAAGCGGGTGAGCAGTTCCTGCGGCG
57400     +---------+---------+---------+---------+---------+---------   57459
          GCTCGCGGAGCTGGCTCCGGGCGGGGTACGGGGCCTTCGCCCACTCGTCAAGGACGCCGC
36          A  L  A  E  V  S  A  R  G  M  G  R  F  R  T  L  L  E  Q  P   -

          GGAAGCCCTTGAACTGCTGCAGAGCCGCGTTGACCGCTCCGTCGAAGTTGCCTGCCGCGG
57460     +---------+---------+---------+---------+---------+---------   57519
          CCTTCGGGAACTTGACGACGTCTCGGCGCAACTGGCGAGGCAGCTTCAACGGACGGCGCC
36          P  F  G  K  F  Q  Q  L  A  A  N  V  A  G  D  F  N  G  A  A   -

          CGGCCGTGACGGCCTTGAACTCCTCCGGCGCCGCCGCCGCGGTCCTGGCCAGGTTGAAGA
57520     +---------+---------+---------+---------+---------+---------   57579
          GCCGGCACTGCCGGAACTTGAGGAGGCCGCGGCGGCGGCGCCAGGACCGGTCCAACTTCT
36          A  A  T  V  A  K  F  E  E  P  A  A  A  A  T  R  A  L  N  F   -

          CGAAGGTGAGGTGGGCGATCTGGTCGGTGACGGTCCAGCCGGGCGCCGGCGTCGGAGTGT
57580     +---------+---------+---------+---------+---------+---------   57639
          GCTTCCACTCCACCCGCTAGACCAGCCACTGCCAGGTCGGCCCGCGGCCGCAGCCTCACA
36          V  F  T  L  H  A  I  Q  D  T  V  T  W  G  P  A  P  T  P  T   -

          TCCAGGCTTCGTCGTCGATCTTCTCGACCAGCTGCGCCAGCTCCTCGATGTCGGTGGCCA
57640     +---------+---------+---------+---------+---------+---------   57699
          AGGTCCGAAGCAGCAGCTAGAAGAGCTGGTCGACGCGGTCGAGGAGCTACAGCCACCGGT
36          N  W  A  E  D  D  I  K  E  V  L  Q  A  L  E  E  I  D  T  A   -

          GGTGCTTGAGGACGTCGTCGAGCGAATTCATCTCGTACTTCCTTCACTGGGGGTGTTCCG
57700     +---------+---------+---------+---------+---------+---------   57759
          CCACGAACTCCTGCAGCAGCTCGCTTAAGTAGAGCATGAAGGAAGTGACCCCCACAAGGC
36          L  H  K  L  V  D  D  L  S  N  M  (ORF36)

          GGCTGGGACGGATGTCCCGCCGGGTGGGCCGGCGGCCGGCGGAAGCGCCGTCGCGGAGCG
57760     +---------+---------+---------+---------+---------+---------   57819
          CCGACCCTGCCTACAGGGCGGCCCACCCGGCCGCCGGCCGCCTTCGCGGCAGCGCCTCGC

          TCGGCGACAGTCGCTAGGCGGCGCGTCCCGCGTAGGAGCCGGCCCGGTCGGAATAGGGCG
57820     +---------+---------+---------+---------+---------+---------   57879
          AGCCGCTGTCAGCGATCCGCCGCGCAGGGCGCATCCTCGGCCGGGCCAGCCTTATCCCGC
37          (ORF37)  *  A  A  R  G  A  Y  S  G  A  R  D  S  Y  P

          CGAGCGCCTCGGCCAGGGCTTCGGGTATCAGGGTCGGCACGGTCGCCGTGTTGGGGCCGC
```

72

```
                                ●                        ●

     GCTCGCGGAGCCGGTCCCGAAGCCCATAGTCCCAGCCGTGCCAGCGGCACAACCCCGGCG
57880 +---------+---------+---------+---------+---------+--------- 57939
     CGAGCGCCTCGGCCAGGGCTTCGGGTATCAGGGTCGGCACGGTCGCCGTGTTGGGGCCGC
37     A  L  A  E  A  L  A  E  P  I  L  T  P  V  T  A  T  N  P  G   -

     GCATGCAGGCGATGCGCTGGCGTCCCCGCGCCACCAGGGTCTCGCCGCCGTCGTCGCCCA
57940 +---------+---------+---------+---------+---------+--------- 57999
     CGTACGTCCGCTACGCGACCGCAGGGGCGCGGTGGTCCCAGAGCGGCGGCAGCAGCGGGT
37     R  M  C  A  I  R  Q  R  G  R  A  V  L  T  E  G  G  D  D  G   -

     GCTTGATGTAGTCGAAGGTGAACTCCAGCTGGGTCTGCCGCAGCTCCGAGAGCCTCATCC
58000 +---------+---------+---------+---------+---------+--------- 58059
     CGAACTACATCAGCTTCCACTTGAGGTCGACCCAGACGGCGTCGAGGCTCTCGGAGTAGG
37     L  K  I  Y  D  F  T  F  E  L  Q  T  Q  R  L  E  S  L  R  M   -

     GGATCGACAGTTCGTCGAAGGCGGTGATCTCCGCGAAGAACTCGCAGTCCACCTTGAGGG
58060 +---------+---------+---------+---------+---------+--------- 58119
     CCTAGCTGTCAAGCAGCTTCCGCCACTAGAGGCGCTTCTTGAGCGTCAGGTGGAACTCCC
37     R  I  S  L  E  D  F  A  T  I  E  A  F  F  E  C  D  V  K  L   -

     TGAAGAGCTTGAGGTCCTCCTGGACCTCGGCGAGCACCGAAGGCGCCCTCTCCTTGAGAA
58120 +---------+---------+---------+---------+---------+--------- 58179
     ACTTCTCGAACTCCAGGAGGACCTGGAGCCGCTCGTGGCTTCCGCGGGAGAGGAACTCTT
37     T  F  L  K  L  D  E  Q  V  E  A  L  V  S  P  A  R  E  K  L   -

     AGAGTTCCCGGCAACGCCCCTGCCAACGAAGGTAGTTGACGTAGTAGACGTTGCCGACGA
58180 +---------+---------+---------+---------+---------+--------- 58239
     TCTCAAGGGCCGTTGCGGGGACGGTTGCTTCCATCAACTGCATCATCTGCAACGGCTGCT
37     F  L  E  R  C  R  G  Q  W  R  L  Y  N  V  Y  Y  V  N  G  V   -

     GGTTCGTCTCCTCGAAGCCGACGGTGTGGCGGAGCTCGAAGTAGTCAGGATTCGTCGCGG
58240 +---------+---------+---------+---------+---------+--------- 58299
     CCAAGCAGAGGAGCTTCGGCTGCCACACCGCCTCGAGCTTCATCAGTCCTAAGCAGCGCC
37     L  N  T  E  E  F  G  V  T  H  R  L  E  F  Y  D  P  N  T  A   -

     TCATAGGTCTGTGCCCTTCGTCGTCGGGGCCGGTCGTCGCACCGAGTTGCGTGAAGCAAC
58300 +---------+---------+---------+---------+---------+--------- 58359
     AGTATCCAGACACGGGAAGCAGCAGCCCCGGCCAGCAGCGTGGCTCAACGCACTTCGTTG
37     T  M  (ORF37)

     TCACTGGTCGCGATGGCCTGCGGGGTCGGTGGCCCGCGCTCCGGGCGGAGAGTGCGGGCG
58360 +---------+---------+---------+---------+---------+--------- 58419
     AGTGACCAGCGCTACCGGACGCCCCAGCCACCGGGCGCGAGGCCCGCCTCTCACGCCCGC

     GGGTGCCGGCCGGCGCGGGGTCAGCCGCGCGCCGACGGCAGCAGGGGAAGAACCCTCTCG
58420 +---------+---------+---------+---------+---------+--------- 58479
     CCCACGGCCGGCCGCGCCCCAGTCGGCGCGCGGCTGCCGTCGTCCCCTTCTTGGGAGAGC
38                  (ORF38)  *  G  R  A  S  P  L  L  P  L  V  R  E  -

     CGGCCGCTCGTGGAGCCGTCGGGGGCCGGTGCGCCGTAGGTGACGGAGATACCCCGGCTC
58480 +---------+---------+---------+---------+---------+--------- 58539
     GCCGGCGAGCACCTCGGCAGCCCCCGGCCACGCGGCATCCACTGCCTCTATGGGGCCGAG
38     R  G  S  T  S  G  D  P  A  P  A  G  Y  T  V  S  I  G  R  S   -

     TGCGCGGCGCGCACGATCCCCGGCATCGCGCGTTCGGCGAGCGCCGCGATGGTCATCGCG
58540 +---------+---------+---------+---------+---------+--------- 58599
     ACGCGCCGCGCGTGCTAGGGGCCGTAGCGCGCAAGCCGCTCGCGGCGCTACCAGTAGCGC
38     Q  A  A  R  V  I  G  P  M  A  R  E  A  L  A  A  I  T  M  A   -

     GGATTGACCGTCAGCGCGCCGGGAACCGACGATCCGTCGGTGACGAAGATCCCCGGGTGG
58600 +---------+---------+---------+---------+---------+--------- 58659
     CCTAACTGGCAGTCGCGCGGCCCTTGGCTGCTAGGCAGCCACTGCTTCTAGGGGCCCACC
38     P  N  V  T  L  A  G  P  V  S  S  G  D  T  V  F  I  G  P  H   -

     TCGCGGAGCTCGTTGCTGTCGTCCAGGGCGGATGTGTGGGGGGTCGTCGCCCATCCGGCAG
58660 +---------+---------+---------+---------+---------+--------- 58719
     AGCGCCTCGAGCAACGACAGCAGGTCCCGCCTACACACCCCCAGCAGCGGGTAGGCCGTC
```

73

```
 38         D   R   L   E   N   S   D   D   L   A   S   T   H   P   D   D   G   M   R   C   -

           GAGGAGAGCGGGTGGACGGTGTAGGCGCCGACGAGGTCGTTGGTCCAGGGCATGACCTTG
    58720  +---------+---------+---------+---------+---------+---------  58779
           CTCCTCTCGCCCACCTGCCACATCCGCGGCTGCTCCAGCAACCAGGTCCCGTACTGGAAC
 38          S   S   L   P   H   V   T   Y   A   G   V   L   D   N   T   W   P   M   V   K   -

           GCCAGGCCGTCCTTCTCCAGGATCTCCTTGACCTCGGCGTCGGATGCGGCCCAGGCGCCC
    58780  +---------+---------+---------+---------+---------+---------  58839
           CGGTCCGGCAGGAAGAGGTCCTAGAGGAACTGGAGCCGCAGCCTACGCCGGGTCCGCGGG
 38          A   L   G   D   K   E   L   I   E   K   V   E   A   D   S   A   A   W   A   G   -

           AGGGTGTTCTTCGTCGGGTCGTAGCGCAGGTTGCCCCGGCCGAGCATCTGCTGGGAGATG
    58840  +---------+---------+---------+---------+---------+---------  58899
           TCCCACAAGAAGCAGCCCAGCATCGCGTCCAACGGGGCCGGCTCGTAGACGACCCTCTAC
 38          L   T   N   K   T   P   D   Y   R   L   N   G   R   G   L   M   Q   Q   S   I   -

           CGGTGGGCGTTACCGGTGGCGGGAGGGGGGCCGAAGACGCCTTCGTTGTCGTCCTCGATC
    58900  +---------+---------+---------+---------+---------+---------  58959
           GCCACCCGCAATGGCCACCGCCCTCCCCCCGGCTTCTGCGGAAGCAACAGCAGGAGCTAG
 38          R   H   A   N   G   T   A   P   P   P   G   F   V   G   E   N   D   D   E   I   -

           ATCGTGAAGATCGTGAGCCAGGAGGTCCACTGCTTCAGGATCTCCTTCTTCTCCTTGCCG
    58960  +---------+---------+---------+---------+---------+---------  59019
           TAGCACTTCTAGCACTCGGTCCTCCAGGTGACGAAGTCCTAGAGGAAGAAGAGGAACGGC
 38          M   T   F   I   T   L   W   S   T   W   Q   K   L   I   E   K   K   E   K   G   -

           AACCAGGAGGGGCCCGTGGCGCCGGGCACCTGGGCGAGGATCGTGCCGAGGCCCGGCGGG
    59020  +---------+---------+---------+---------+---------+---------  59079
           TTGGTCCTCCCCGGGCACCGCGGCCCGTGGACCCGCTCCTAGCACGGCTCCGGGCCGCCC
 38          F   W   S   P   G   T   A   G   P   V   Q   A   L   I   T   G   L   G   P   P   -

           AAGTAGAGCTGTTCCAGGGAGTAGCGGGAGTACTCGGGCAACGAGCCGTCCAGCCTGTCC
    59080  +---------+---------+---------+---------+---------+---------  59139
           TTCATCTCGACAAGGTCCCTCATCGCCCTCATGAGCCCGTTGCTCGGCAGGTCGGACAGG
 38          F   Y   L   Q   E   L   S   Y   R   S   Y   E   P   L   S   G   D   L   R   D   -

           CAGCTCGCCACGGTGGGCCCCTTGCCGATCTGGTTGGCCGCGTAGGCGAGCCCGTCGCCC
    59140  +---------+---------+---------+---------+---------+---------  59199
           GTCGAGCGGTGCCACCCGGGGAACGGCTAGACCAACCGGCGCATCCGCTCGGGCAGCGGG
 38          W   S   A   V   T   P   G   K   G   I   Q   N   A   A   Y   A   L   G   D   G   -

           CGGTCCAGGCCGAACAGCTCGGCCGCCTTGGCCTCGTCGATGATGGCGGTGTTGAGCCGC
    59200  +---------+---------+---------+---------+---------+---------  59259
           GCCAGGTCCGGCTTGTCGAGCCGGCGGAACCGGAGCAGCTACTACCGCCACAACTCGGCG
 38          R   D   L   G   F   L   E   A   A   K   A   E   D   I   I   A   T   N   L   R   -

           TCGCCGTTGCCGGAGAAGTAGCGTCCGACCGCTCGTGGCATGGTGCCCAGGTGGGCCTCG
    59260  +---------+---------+---------+---------+---------+---------  59319
           AGCGGCAACGGCCTCTTCATCGCAGGCTGGCGAGCACCGTACCACGGGTCCACCCGGAGC
 38          E   G   N   G   S   F   Y   R   G   V   A   R   P   M   T   G   L   H   A   E   -

           CTGCGCTGGAGGATCACCGGGGGTCGCGCCCGCGCCGGCCGCCATCACCACGATCTTCGCC
    59320  +---------+---------+---------+---------+---------+---------  59379
           GACGCGACCTCCTAGTGGCCCCAGCGCGGGCGCGGCCGGCGGTAGTGGTGCTAGAAGCGG
 38          S   R   Q   L   I   V   P   T   A   G   A   G   A   A   M   V   V   I   K   A   -

           TCGATGACGCCGCTGCCCGCCTGGAGGCGGTAGTCGTCGTCGTGCACGACGTTGTAGTGC
    59380  +---------+---------+---------+---------+---------+---------  59439
           AGCTACTGCGGCGACGGGCGGACCTCCGCCATCAGCAGCAGCACGTGCTGCAACATCACG
 38          E   I   V   G   S   G   A   Q   L   R   Y   D   D   D   H   V   V   N   Y   H   -

           ACCCGGTAGGAGCCGTCGGGGGTGCGCGAGAGGTGCTGGACCTCGTGCAGCGGGCGGATG
    59440  +---------+---------+---------+---------+---------+---------  59499
           TGGGCCATCCTCGGCAGCCCCCACGCGCTCTCCACGACCTGGAGCACGTCGCCCGCCTAC
 38          V   R   Y   S   G   D   P   T   R   S   L   H   Q   V   E   H   L   P   R   I   -
```

74

```
               CGCGCCCCATGGGCGATGGCGGCGGGCAGGTAGTTGACCAGCAAGGACTGCTTGGCCTCG
     59500     +---------+---------+---------+---------+---------+---------  59559
               GCGCGGGGTACCCGCTACCGCCGCCCGTCCATCAACTGGTCGTTCCTGACGAACCGGAGC
     38         R  A  G  H  A  I  A  A  P  L  Y  N  V  L  L  S  Q  K  A  E  -

               AAGCGGCAGCCGGCCATCATCCAGTTGCAGTTCACGCACTTGGTGTTGTCGATGGCGACG
     59560     +---------+---------+---------+---------+---------+---------  59619
               TTCGCCGTCGGCCGGTAGTAGGTCAACGTCAAGTGCGTGAACCACAACAGCTACCGCTGC
     38         F  R  C  G  A  M  M  W  N  C  N  V  C  K  T  N  D  I  A  V  -

               GCGAGGGGGTTGGCGGTGCGGCCGGCGTGGTTGCACGCCGCGGCCCACAGTCCGCCGGCG
     59620     +---------+---------+---------+---------+---------+---------  59679
               CGCTCCCCCAACCGCCACGCCGGCCGCACCAACGTGCGGCGCCGGGTGTCAGGCGGCCGC
     38         A  L  P  N  A  T  R  G  A  H  N  C  A  A  A  W  L  G  G  A  -

               TAGCTCACGTCGTTCCAGTCCTGCCGGGTCACGGAGAGGGACTCCTCGACACGGTCGTAC
     59680     +---------+---------+---------+---------+---------+---------  59739
               ATCGAGTGCAGCAAGGTCAGGACGGCCCAGTGCCTCTCCCTGAGGAGCTGTGCCAGCATG
     38         Y  S  V  D  N  W  D  Q  R  T  V  S  L  S  E  E  V  R  D  Y  -

               CAGGGGTCCAGGGTTTCGCGGCTCACCGCCTGCGGCCACATCCGGCGTCCTATGGACCCC
     59740     +---------+---------+---------+---------+---------+---------  59799
               GTCCCCAGGTCCCAAAGCGCCGAGTGGCGGACGCCGGTGTAGGCCGCAGGATACCTGGGG
     38         W  P  D  L  T  E  R  S  V  A  Q  P  W  M  R  R  G  I  S  G  -

               TGCCGGTCGAAGACGAAGCGCGGGGCGCGGGGCATCGCGGCGAAGTAGACGACGCTGCCG
     59800     +---------+---------+---------+---------+---------+---------  59859
               ACGGCCAGCTTCTGCTTCGCGCCCCGCGCCCCGTAGCGCCGCTTCATCTGCTGCGACGGC
     38         Q  R  D  F  V  F  R  P  A  R  P  M  A  A  F  Y  V  V  S  G  -

               CCGCCCACACAGTTCCCGCCGAGGATGCTCATGCCGTCCCCGACCGTGAAGTCGAACGCC
     59860     +---------+---------+---------+---------+---------+---------  59919
               GGCGGGTGTGTCAAGGGCGGCTCCTACGAGTACGGCAGGGGCTGGCACTTCAGCTTGCGG
     38         G  G  V  C  N  G  G  L  I  S  M  G  D  G  V  T  F  D  F  A  -

               CTCGTGTACGAGGAGCCGAGTTTGTAGTCGTGCTCGAACTCCTTGCTCTCCAGCCACGGC
     59920     +---------+---------+---------+---------+---------+---------  59979
               GAGCACATGCTCCTCGGCTCAAACATCAGCACGAGCTTGAGGAACGAGAGGTCGGTGCCG
     38         R  T  Y  S  S  G  L  K  Y  D  H  E  F  E  K  S  E  L  W  P  -

               CCGCGTTCCAGGACGGTGACGTCGGCGCCCCCCGCCGCCAGGTGGTAGGCGGCGATGGCA
     59980     +---------+---------+---------+---------+---------+---------  60039
               GGCGCAAGGTCCTGCCACTGCAGCCGCGGGGGGCGGCGGTCCACCATCCGCCGCTACCGT
     38         G  R  E  L  V  T  V  D  A  G  G  A  A  L  H  Y  A  A  I  A  -

               CCGCCGAATCCGCTGCCGATGACGAGGACGTCCGTGCGCTCGGCCGTGGTGCTCATGCGG
     60040     +---------+---------+---------+---------+---------+---------  60099
               GGCGGCTTAGGCGACGGCTACTGCTCCTGCAGGCACGCGAGCCGGCACCACGAGTACGCC
     38         G  G  F  G  S  G  I  V  L  V  D  T  R  E  A  T  T  S  M
                                                          (ORF39)  *  A   -

               GGCTCCCGGTGGACGTGGTGTCGGGGTGGAGGCGGGCGAACTCACGCCCGTAGCTGTAAT
     60100     +---------+---------+---------+---------+---------+---------  60159
               CCGAGGGCCACCTGCACCACAGCCCCACCTCCGCCCGCTTGAGTGCGGGCATCGACATTA
     39         P  S  G  T  S  T  T  D  P  H  L  R  A  F  E  R  G  Y  S  Y  -

               CCTTGAAGCGCCACAGGCCGTCGGCGTCCGGCATGCTCAGGCCCATGGCCTCCAGTCCCG
     60160     +---------+---------+---------+---------+---------+---------  60219
               GGAACTTCGCGGTGTCCGGCAGCCGCAGGCCGTACGAGTCCGGGTACCGGAGGTCAGGGC
     39         D  K  F  R  W  L  G  D  A  D  P  M  S  L  G  M  A  E  L  G  -

               GATGGCCGTCCTCCATCGCCTGTGCCGTGTTGAGGTGCGCGGCCGAATCGAAGGCCATGT
     60220     +---------+---------+---------+---------+---------+---------  60279
               CTACCGGCAGGAGGTAGCGGACACGGCACAACTCCACGCGCCGGCTTAGCTTCCGGTACA
     39         P  H  G  D  E  M  A  Q  A  T  N  L  H  A  A  S  D  F  A  M  -
```

75

```
                  TGCAGAAGAGGGACAGCAGCACCCAGAACTCCTTCTCGGGGTGGCCTGGTGTCGTCAGCC
         60280    +---------+---------+---------+---------+---------+---------  60339
                  ACGTCTTCTCCCTGTCGTCGTGGGTCTTGAGGAAGAGCCCCACCGGACCACAGCAGTCGG
         39         N  C  F  L  S  L  L  V  W  F  E  K  E  P  H  G  P  T  T  L   -

                  GCTGGATCAGCGCGGCCCGGTCCGGGTAGTCGAGCGCCACGAAGGGCGGGACCGTCGGGT
         60340    +---------+---------+---------+---------+---------+---------  60399
                  CGACCTAGTCGCGCCGGGCCAGGCCCATCAGCTCGCGGTGCTTCCCGCCCTGGCAGCCCA
         39         R  Q  I  L  A  A  R  D  P  Y  D  L  A  V  F  P  P  V  T  P   -

                  CGGGAGCCAGGCGGCGCTCCGCCGCGTAGGCCAGCGCGTGCTCGTTCACCAGGCGCACCA
         60400    +---------+---------+---------+---------+---------+---------  60459
                  GCCCTCGGTCCGCCGCGAGGCGGCGCATCCGGTCGCGCACGAGCAAGTGGTCCGCGTGGT
         39         D  P  A  L  R  R  E  A  A  Y  A  L  A  H  E  N  V  L  R  V   -

                  GGTCGTCCAGACCCTCGTGGATGCCGGTCGCATCCCATTGCAGGAGCTCCAGGGCTCCCG
         60460    +---------+---------+---------+---------+---------+---------  60519
                  CCAGCAGGTCTGGGAGCACCTACGGCCAGCGTAGGGTAACGTCCTCGAGGTCCCGAGGGC
         39         L  D  D  L  G  E  H  I  G  T  A  D  W  Q  L  L  E  L  A  G   -

                  CCTGGACGGCGCCACCGCCGGTGGACACCCCCGCGATGGCCCGGTCGTCCGCGAAGCGCT
         60520    +---------+---------+---------+---------+---------+---------  60579
                  GGACCTGCCGCGGTGGCGGCCACCTGTGGGGGCGCTACCGGGCCAGCAGGCGCTTCGCGA
         39         A  Q  V  A  G  G  G  T  S  V  G  A  I  A  R  D  D  A  F  R   -

                  TCTGGCCCGGCACGATCGTGTCCGCGTAGGCCTCCAGGGTCATGGTCCGGATATCGCCGG
         60580    +---------+---------+---------+---------+---------+---------  60639
                  AGACCGGGCCGTGCTAGCACAGGCGCATCCGGAGGTCCCAGTACCAGGCCTATAGCGGCC
         39         K  Q  G  P  V  I  T  D  A  Y  A  E  L  T  M (ORF39)

                  CCGGCGCCCCTCGCTCATTGTCGTCGCGCAACTCGCTCTCCATTCTCGCAGTCCGGAGTG
         60640    +---------+---------+---------+---------+---------+---------  60699
                  GGCCGCGGGGAGCGAGTAACAGCAGCGCGTTGAGCGAGAGGTAAGAGCGTCAGGCCTCAC

                  GGATGCCTTGTGGCGAGGAGAAAGCTAGGTTCGTTCGACCGGTTCAAGCAACTAGCCAAA
         60700    +---------+---------+---------+---------+---------+---------  60759
                  CCTACGGAACACCGCTCCTCTTTCGATCCAAGCAAGCTGGCCAAGTTCGTTGATCGGTTT

                  GTCGAGGCGACCTTGAAACCGACTCCACGGAGTTGGCGCGAAGCGGCGGATGGATTACAC
         60760    +---------+---------+---------+---------+---------+---------  60819
                  CAGCTCCGCTGGAACTTTGGCTGAGGTGCCTCAACCGCGCTTCGCCGCCTACCTAATGTG

                  GCGCGGGCGAGCGGCTCACTAGTCTGGCCGCACGGATGTCTTCATCACCTGCACGTGGAA
         60820    +---------+---------+---------+---------+---------+---------  60879
                  CGCGCCCGCTCGCCGAGTGATCAGACCGGCGTGCCTACAGAAGTAGTGGACGTGCACCTT

                  AAGCTTCTGCACGGGCACCGCATGTGGAAGTGAGCCCTGGTCTCATGTCTTGGGGGAAAC
         60880    +---------+---------+---------+---------+---------+---------  60939
                  TTCGAAGACGTGCCCGTGGCGTACACCTTCACTCGGGACCAGAGTACAGAACCCCCTTTG

                  GTGAAAAGTGACTCTGCCCAACGCGCCGTGGAGCGATCACGCCGTGTCGTACGGATCGAT
         60940    +---------+---------+---------+---------+---------+---------  60999
                  CACTTTTCACTGAGACGGGTTGCGCGGCACCTCGCTAGTGCGGCACAGCATGCCTAGCTA
         40         V  K  S  D  S  A  Q  R  A  V  E  R  S  R  R  V  V  R  I  D   -

                  (ORF40)
                  GAACTCATTCCCGCCGATTCCCCGCGCCTGAACGGAATCGATCGTTCCCATGTGCAGCGC
         61000    +---------+---------+---------+---------+---------+---------  61059
                  CTTGAGTAAGGGCGGCTAAGGGGCGCGGACTTGCCTTAGCTAGCAAGGGTACACGTCGCG
         40         E  L  I  P  A  D  S  P  R  L  N  G  I  D  R  S  H  V  Q  R   -

                  CTCGCGACCGTGTACGCGTCCCTGCCGCCGGTCCTGGTGCACCGCCCGACCATGCGGGTC
         61060    +---------+---------+---------+---------+---------+---------  61119
                  GAGCGCTGGCACATGCGCAGGGACGGCGGCCAGGACCACGTGGCGGGCTGGTACGCCCAG
         40         L  A  T  V  Y  A  S  L  P  P  V  L  V  H  R  P  T  M  R  V   -
```

76

```
                    GTCGACGGCATGCACCGCATCGGCGCGGCCCGCCTGAAGGGGCTGGACACGGTCGAGGTC
        61120       +---------+---------+---------+---------+---------+---------  61179
                    CAGCTGCCGTACGTGGCGTAGCCGCGCCGGGCGGACTTCCCCGACCTGTGCCAGCTCCAG
         40           V  D  G  M  H  R  I  G  A  A  R  L  K  G  L  D  T  V  E  V   -

                    ACCTTCTTCGAGGGCGCCGAGGAGCAGGTGTTCCTGCGTTCCGTCGCGGCGAACATCACC
        61180       +---------+---------+---------+---------+---------+---------  61239
                    TGGAAGAAGCTCCCGCGGCTCCTCGTCCACAAGGACGCAAGGCAGCGCCGCTTGTAGTGG
         40           T  F  F  E  G  A  E  E  Q  V  F  L  R  S  V  A  A  N  I  T   -

                    AACGGCCTGCCGTTGTCGGTGGCCGACCGCAAGACCGCCGCGGCCCGCATTCTGGCCTCC
        61240       +---------+---------+---------+---------+---------+---------  61299
                    TTGCCGGACGGCAACAGCCACCGGCTGGCGTTCTGGCGGCGCCGGGCGTAAGACCGGAGG
         40           N  G  L  P  L  S  V  A  D  R  K  T  A  A  A  R  I  L  A  S   -

                    CACCCGACCCTGTCCGACCGCGCGGTCGCCGCACACGTCGGCCTCGACGCCAAGACCGTG
        61300       +---------+---------+---------+---------+---------+---------  61359
                    GTGGGCTGGGACAGGCTGGCGCGCCAGCGGCGTGTGCAGCCGGAGCTGCGGTTCTGGCAC
         40           H  P  T  L  S  D  R  A  V  A  A  H  V  G  L  D  A  K  T  V   -

                    GCGGGGGTACGGACGTGTTCAGCCGCGGGTTCTCCGCTGCTGAACATGCGCACCGGGGCG
        61360       +---------+---------+---------+---------+---------+---------  61419
                    CGCCCCCATGCCTGCACAAGTCGGCGCCCAAGAGGCGACGACTTGTACGCGTGGCCCCGC
         40           A  G  V  R  T  C  S  A  A  G  S  P  L  L  N  M  R  T  G  A   -

                    GACGGCCGCGTCCACCCGTTGGACCGCACCGCCGAACGCCTGCACGCGGCCGCGCTGCTG
        61420       +---------+---------+---------+---------+---------+---------  61479
                    CTGCCGGCGCAGGTGGGCAACCTGGCGTGGCGGCTTGCGGACGTGCGCCGGCGCGACGAC
         40           D  G  R  V  H  P  L  D  R  T  A  E  R  L  H  A  A  A  L  L   -

                    ACCCAGGACCCGGGACTCCCGTTGCGCTCCGTCGTCGAGCAGACGGGGCTGTCGCTGGGC
        61480       +---------+---------+---------+---------+---------+---------  61539
                    TGGGTCCTGGGCCCTGAGGGCAACGCGAGGCAGCAGCTCGTCTGCCCCGACAGCGACCCG
         40           T  Q  D  P  G  L  P  L  R  S  V  V  E  Q  T  G  L  S  L  G   -

                    ACGGCCCACGACGTCCGCCGTCGGCTGCTGCGGGGCGAGGACCCGGTCCCGCAGAACCGG
        61540       +---------+---------+---------+---------+---------+---------  61599
                    TGCCGGGTGCTGCAGGCGGCAGCCGACGACGCCCCGCTCCTGGGCCAGGGCGTCTTGGCC
         40           T  A  H  D  V  R  R  R  L  L  R  G  E  D  P  V  P  Q  N  R   -

                    CAGAGCGCGATGCTGGAGCCGGGACTCGCCCCGCAGAAGAAGGCGACGGCCAAGCCGCCC
        61600       +---------+---------+---------+---------+---------+---------  61659
                    GTCTCGCGCTACGACCTCGGCCCTGAGCGGGGCGTCTTCTTCCGCTGCCGGTTCGGCGGG
         40           Q  S  A  M  L  E  P  G  L  A  P  Q  K  K  A  T  A  K  P  P   -

                    GTCGGCCCGGCCGCCCGTCCGGTCCCGAAGGTGCCGCCCGCCGTCGCCGGCAGGCCGCCG
        61660       +---------+---------+---------+---------+---------+---------  61719
                    CAGCCGGGCCGGCGGGCAGGCCAGGGCTTCCACGGCGGGCGGCAGCGGCCGTCCGGCGGC
         40           V  G  P  A  A  R  P  V  P  K  V  P  P  A  V  A  G  R  P  P   -

                    GTGTCACCGCGGTCCCGGGCCCCGCTGGAGGCGCTGCGCAAGCTCTCCAACGACCCCTCC
        61720       +---------+---------+---------+---------+---------+---------  61779
                    CACAGTGGCGCCAGGGCCCGGGGCGACCTCCGCGACGCGTTCGAGAGGTTGCTGGGGAGG
         40           V  S  P  R  S  R  A  P  L  E  A  L  R  K  L  S  N  D  P  S   -

                    CTGCGCCACTCCGACCAGGGGCGCGAACTCATGCGCTGGCTGCACAACCGGTTCGTCGTC
        61780       +---------+---------+---------+---------+---------+---------  61839
                    GACGCGGTGAGGCTGGTCCCCGCGCTTGAGTACGCGACCGACGTGTTGGCCAAGCAGCAG
         40           L  R  H  S  D  Q  G  R  E  L  M  R  W  L  H  N  R  F  V  V   -

                    GACGAGGCGTGGCGCCGGCGCGCGGACGCGGTCCCGGCCCACTGCGTCGACTCGATGGCG
        61840       +---------+---------+---------+---------+---------+---------  61899
                    CTGCTCCGCACCGCGGCCGCGCGCCTGCGCCAGGGCCGGGTGACGCAGCTGAGCTACCGC
         40           D  E  A  W  R  R  R  A  D  A  V  P  A  H  C  V  D  S  M  A   -
```

77

```
        GAGCTGGCGCAGCACTGCTCGGACGCCTGGCACCGGTTCGCCGAGGAGATGGTTCGGCGC
  61900 +---------+---------+---------+---------+---------+---------  61959
        CTCGACCGCGTCGTGACGAGCCTGCGGACCGTGGCCAAGCGGCTCCTCTACCAAGCCGCG
   40      E  L  A  Q  H  C  S  D  A  W  H  R  F  A  E  E  M  V  R  R    -

        CGGCACAGCGCCGCGGCCGACGGCTCCGGACTCCGCACGACTCAGCCAACTCGCCGTTGA
  61960 +---------+---------+---------+---------+---------+---------  62019
        GCCGTGTCGCGGCGCCGGCTGCCGAGGCCTGAGGCGTGCTGAGTCGGTTGAGCGGCAACT
   40      R  H  S  A  A  A  D  G  S  G  L  R  T  T  Q  P  T  R  R  *
        (ORF40) -

        CGGCCTACTTCGACAGGGAGTTACGGTGACCACGAACACCATCGAGGACGCGGTCCGCCG
  62020 +---------+---------+---------+---------+---------+---------  62079
        GCCGGATGAAGCTGTCCCTCAATGCCACTGGTGCTTGTGGTAGCTCCTGCGCCAGGCGGC
   41                            (ORF41)    V  T  T  N  T  I  E  D  A  V  R  R

        GGTCGTCGAGTACATGCACGTCAACCTGGGTCAGAACCTCACGATCGATGACATGGCGCG
  62080 +---------+---------+---------+---------+---------+---------  62139
        CCAGCAGCTCATGTACGTGCAGTTGGACCCAGTCTTGGAGTGCTAGCTACTGTACCGCGC
   41        V  V  E  Y  M  H  V  N  L  G  Q  N  L  T  I  D  D  M  A  R

        CACGGCGATGTTCAGCAAGTTCCATTTCACCCGCATCTTCCGCGAAGTCACCGGTACCTC
  62140 +---------+---------+---------+---------+---------+---------  62199
        GTGCCGCTACAAGTCGTTCAAGGTAAAGTGGGCGTAGAAGGCGCTTCAGTGGCCATGGAG
   41      T  A  M  F  S  K  F  H  F  T  R  I  F  R  E  V  T  G  T  S    -

        TCCCGGGCGTTTCCTGTCCGCCTTACGGATTCAGGAGGCCAAGAGACTTCTCGTGCACAC
  62200 +---------+---------+---------+---------+---------+---------  62259
        AGGGCCCGCAAAGGACAGGCGGAATGCCTAAGTCCTCCGGTTCTCTGAAGAGCACGTGTG
   41      P  G  R  F  L  S  A  L  R  I  Q  E  A  K  R  L  L  V  H  T    -

        TGCACTCAGTGTGGCCGATATCAGCAGTCAGGTCGGCTACAGCAGTGTCGGTACTTTCAG
  62260 +---------+---------+---------+---------+---------+---------  62319
        ACGTGAGTCACACCGGCTATAGTCGTCAGTCCAGCCGATGTCGTCACAGCCATGAAAGTC
   41      A  L  S  V  A  D  I  S  S  Q  V  G  Y  S  S  V  G  T  F  S    -

        TTCTCGCTTCAAGGCCTGTGTGGGGCTTTCCCCGAGCGCCTATCGCGACTTCGGCGGGGT
  62320 +---------+---------+---------+---------+---------+---------  62379
        AAGAGCGAAGTTCCGGACACACCCCGAAAGGGGCTCGCGGATAGCGCTGAAGCCGCCCCA
   41      S  R  F  K  A  C  V  G  L  S  P  S  A  Y  R  D  F  G  G  V    -

        GCAGCCGGGTTTTCCCTCCGCCGCGGCCCGTCTCACTCCCACCGCGCACAATCCCTCCGT
  62380 +---------+---------+---------+---------+---------+---------  62439
        CGTCGGCCCAAAAGGGAGGCGGCGCCGGGCAGAGTGAGGGTGGCGCGTGTTAGGGAGGCA
   41      Q  P  G  F  P  S  A  A  A  R  L  T  P  T  A  H  N  P  S  V    -

        GCGCGGCCGCATTCACTCCGCCCCGGGTGACAGGCCCGGAAGGATCTTCGTGGGCCTGTT
  62440 +---------+---------+---------+---------+---------+---------  62499
        CGCGCCGGCGTAAGTGAGGCGGGGCCCACTGTCCGGGCCTTCCTAGAAGCACCCGGACAA
   41      R  G  R  I  H  S  A  P  G  D  R  P  G  R  I  F  V  G  L  F    -

        CCCCGGCAGGATGCGCCAGGGCCGCCCGGCGCGCTGGACCGTCATGGAGAGTCCCGGGGC
  62500 +---------+---------+---------+---------+---------+---------  62559
        GGGGCCGTCCTACGCGGTCCCGGCGGGCCGCGCGACCTGGCAGTACCTCTCAGGGCCCCG
   41      P  G  R  M  R  Q  G  R  P  A  R  W  T  V  M  E  S  P  G  A    -

        CTTCGAGCTCCGGGACGTGCCCGTGGGCACCTGGCACATCCTGGTCCACTCCTTCCCCGC
  62560 +---------+---------+---------+---------+---------+---------  62619
        GAAGCTCGAGGCCCTGCACGGGCACCCGTGGACCGTGTAGGACCAGGTGAGGAAGGGGCG
   41      F  E  L  R  D  V  P  V  G  T  W  H  I  L  V  H  S  F  P  A    -

        CGGACACCGGCCGCACCAGCTCGACTCCGAACCGCTGTTGCTCGGGCACAGCGGACCGCT
  62620 +---------+---------+---------+---------+---------+---------  62679
        GCCTGTGGCCGGCGTGGTCGAGCTGAGGCTTGGCGACAACGAGCCCGTGTCGCCTGGCGA
   41      G  H  R  P  H  Q  L  D  S  E  P  L  L  L  G  H  S  G  P  L    -
```

78

```
                CGTGGTGCACCCCGGTGCCCTGCTCCGGCCGGCGGACATCCTCCTGCGCGCGGTGGACGC
         62680  +---------+---------+---------+---------+---------+--------- 62739
                GCACCACGTGGGGCCACGGGACGAGGCCGGCCGCCTGTAGGAGGACGCGCGCCACCTGCG
         41        V  V  H  P  G  A  L  L  R  P  A  D  I  L  L  R  A  V  D  A  -

                CCTCGATCCACCGGTCCTGCTGGCCCACTTCGCGCTGGAGAGCCGCCTCACCTCGCCGTA
         62740  +---------+---------+---------+---------+---------+--------- 62799
                GGAGCTAGGTGGCCAGGACGACCGGGTGAAGCGCGACCTCTCGGCGGAGTGGAGCGGCAT
         41       L  D  P  P  V  L  L  A  H  F  A  L  E  S  R  L  T  S  P  Y  -
         42                                        (ORF42)    *  R  A  T  -

                CTCACCGTCATCGGTAGCCCTCCGCGCATCCGCAGGGAGAGCATGGGTTCGGCAACCGCC
         62800  +---------+---------+---------+---------+---------+--------- 62859
                GAGTGGCAGTAGCCATCGGGAGGCGCGTAGGCGTCCCTCTCGTACCCAAGCCGTTGGCGG
         41        S  P  S  S  V  A  L  R  A  S  A  G  R  A  W  V  R  Q  P  P  -
         42        S  V  T  M  P  L  G  G  R  M  R  L  S  L  M  P  E  A  V  A  -

                CGGTGTCCGGCGACGGTACGCAGATCGAGATCGCGGGTGACCAGGGCCGTGACGAACACC
         62860  +---------+---------+---------+---------+---------+--------- 62919
                GCCACAGGCCGCTGCCATGCGTCTAGCTCTAGCGCCCACTGGTCCCGGCACTGCTTGTGG
         41        G  V  R  R  R  Y  A  D  R  D  R  G  *  (ORF41)
         42        R  H  G  A  V  T  R  L  D  L  D  R  T  V  L  A  T  V  F  V  -

                GCCTCCATCATCCCGAGGTTGCTGCCGACGCAGAACCGGGGCCCCGCGCCGAACGGGATG
         62920  +---------+---------+---------+---------+---------+--------- 62979
                CGGAGGTAGTAGGGCTCCAACGACGGCTGCGTCTTGGCCCCGGGGCGCGGCTTGCCCTAC
         42        A  E  M  M  G  L  N  S  G  V  C  F  R  P  G  A  G  F  P  I  -

                TACGCGTACCGCGGCCGGTCGGCGGTCTGCCGGGGGTTCGAACCGCTCGGGGTCGAAGCGC
         62980  +---------+---------+---------+---------+---------+--------- 63039
                ATGCGCATGGCGCCGGCCAGCCGCCAGACGGCCCCAAGCTTGGCGAGCCCCAGCTTCGCG
         42        Y  A  Y  R  P  R  D  A  T  Q  R  P  E  F  R  E  P  D  F  R  -

                TCGGGGTCCTCCCACAGCCCCGGATGGCGGTGCATGATGTACGGGCAGACCAGCACATCC
         63040  +---------+---------+---------+---------+---------+--------- 63099
                AGCCCCAGGAGGGTGTCGGGGCCTACCGCCACGTACTACATGCCCGTCTGGTCGTGTAGG
         42        E  P  D  E  W  L  G  P  H  R  H  M  I  Y  P  C  V  L  V  D  -

                GATCCGGCGGACACCGTGTAGCCGCCGACCACATCGCGTTGCTGGGCCACCCTGGGCAGG
         63100  +---------+---------+---------+---------+---------+--------- 63159
                CTAGGCCGCCTGTGGCACATCGGCGGCTGGTGTAGCGCAACGACCCGGTGGGACCCGTCC
         42        S  G  A  S  V  T  Y  G  G  V  V  D  R  Q  Q  A  V  R  P  L  -

                ATCCC
         63160  +---- 63164
                TAGGG
         42        I  G  -
```

79